

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2004, 13:18:28 ; Search time 4477.84 Seconds
(without alignments)
9887.411 Million cell updates/sec

Title: US-10-627-124-1

Perfect score: 1215

Sequence: 1 atgagccatcgatcaatt.....tcgtgcgttgacggggca 1215

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: gb_est1: *
2: gb_est2: *
3: gb_hic: *
4: gb_est3: *
5: gb_est4: *
6: gb_est5: *
7: gb_est6: *
8: gb_gsa1: *
9: gb_gsa2: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54.2	4.5	975	8	BZ675691 PUBDH78TD
2	45	3.7	711	5	BW248443 BW248443
3	44.2	3.6	697	2	BE585802 Est#2pT7
4	44	3.6	569	6	CB827487 LjNEST75d
5	42	3.5	1101	9	AL063921 Drosophil
6	41.8	3.4	439	1	AL823229 AL823229
7	41.6	3.4	330	8	AY174775 AY174775
8	40.6	3.3	631	7	CK535040 tswgb0.00
9	40.6	3.3	777	6	CA923083 EST640801
10	40	3.3	747	7	CF347163 AGENCOURT
11	40	3.3	905	7	CN317145 AGENCOURT
12	40	3.3	930	6	CF265727 AGENCOURT
13	39.4	3.2	779	5	BU796025 SJF2DXG07
14	39.2	3.2	585	2	BE432743 EST39272
15	38.8	3.2	705	5	BQ853232 QGB1905.Y
16	38.8	3.2	732	5	BU013226 QGJ4B04.Y
17	38.6	3.2	643	7	CN086902 EC2BBA27D
18	38.6	3.2	662	7	CF388973 RTDR2.16
19	38.6	3.2	670	7	CF477876 RTW3.13
20	38.6	3.2	682	7	CO369247 RTK1.45H
21	38.6	3.2	688	7	CN086901 EC2BBA27D
22	38.6	3.2	689	7	CF479071 RTW3.21
23	38.6	3.2	712	6	CF113752 Shultzomi
24	38.6	3.2	714	7	CR420363 CR420363

ALIGNMENTS

RESULT 1
BZ675691
LOCUS PUBDH78TD ZM 0.6_1.0_KB Zea mays genomic clone ZMMBTA029M12,
DEFINITION genomic survey sequence.
ACCESSION BZ675691
VERSION BZ675691.1 GI:28226790
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 975)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J., and
Bennetzen,J.
TITLE Maize Genomics Consortium
JOURNAL Unpublished (2003)
COMMENT Contact: Cathy Whitelaw
TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
FEATURES
source location/Qualifiers
1..975
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBTA029M12"
/clone_lib="ZM 0.6_1.0_KB"
/note="Vector: pCR4-TOPO; Site:1: EcoRI; 0.6-1.0 kb high
CoT selected genomic DNA library"

ORIGIN

Query Match 4.5%; Score 54.2; DB 8; Length 975;
Best Local Similarity 46.2%; Pred. No. 5.4e-05;
Matches 222; Conservative 0; Mismatches 253; Indels 6; Gaps 1;
QY 514 CCTGAAGCGCGGCACATTCGTTATGTCACGACAGCTTCCCAATCTGATGCGG 573
DB 125 CCTGAAGCGCGGCATCAGACTCAGCTCATCTGCGGAGCTGCGCATACATCAAAAAC 184

QY 574 GTCAATTAAGAAACGTCGCTCAATCCGGATCAGATTTAAATTCGATCCTATGTACTTCT 633
DB 185 CTGGTGGCGAAGCTGCAGCGCAACCAACGCGATGACTTCATGACCGCTTGACCAAGCC 244
QY 634 GAATATGAAGGCATGGCGGTGTCGGACAAGGATATACCTGCACCTGATCTTAAATGTGCTG 693
DB 245 GACGAAAAGGCGAGTCCCTCAACGAGGAGCAATTGCTGCTGTGTGCAATTCGATCCTG 304
QY 694 TTAGCCGCAACGGAACCGGCTGATAGACGCTGGCACTGATCATCTACATTTGCTCAAC 753
DB 305 GTCGCGGTTTGAAGAGGAGTGCCTCGCAGATCCCAACTTATCCAGGTGCTGTGGAC 364
QY 754 AATCTGTAGCAGATGAATGATGTTTGGCTGACCGTTCGTTAGTTCGAGAGCCATTGGG 813
DB 365 AACCCGCGCATGCGACAGCTCAAGCGCGATCCGAGCAGATACCGCAGCGGTGCA 424
QY 814 GAGACATTGCGTTATAAACCCG-----CGGTTACGCTGATTCGCGGCGAGCTGTCCCA 867
DB 425 GAGTGTGCTGCGTACATCCCGCTGCGCTCGCGGCGATGTTTGTGCACTACGCGCTGGAA 484
QY 868 GATACAGTGGTGGCGGTATGAAATCAAAAAGATACAGATTGTTTTTGTATGATCGGT 927
DB 485 GACATTCAGTCCGTGGAACCCCTGTGTGCGCCAGGCGACGCGGTGTTCGCTCGCATCGGC 544
QY 928 GCGGCTAACCGGACCGCTGAAGCATTTGAACAGCCTGACGTGTTTAATATTCATCGGAA 987
DB 545 GCGGCNACCATGACCGCGCCGCTTTGAAAACCCACAGCGCTGGACCTGCAACGGAT 604
QY 988 G 988
DB 605 G 605

RESULT 2

BW248443
LOCUS BW248443 Nori Satoh unpublished cDNA library, EST 09-NOV-2002
DEFINITION intestinalis cDNA clone citb07m23 5', mRNA sequence.

ACCESSION BW248443.1 GI:24828361

VERSION EST.

KEYWORDS

SOURCE

ORGANISM

Ciona intestinalis

Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;

Phlebobranchia; Clonidae; Ciona.

1 (bases 1 to 711)

Satou,Y., Shin-i,T., Kohara,Y. and Satoh,N.

Expressed genes in Ciona intestinalis (2002c)

Unpublished (2002)

Contact: Nori Satoh

Department of Zoology

Kyoto University

Sakyo-ku, Kyoto, Kyoto 606-8502, Japan

Tel: 81-75-753-4081

Fax: 81-75-705-1113

Email: satohe@acidian.zool.kyoto-u.ac.jp.

Location/Qualifiers

1. .711

/organism="Ciona intestinalis"

/mol_type="mRNA"

/db_xref="taxon:7719"

/clone="citb07m23"

/tissue_type="whole animal"

/dev_stage="tailbud embryo"

/clone_lib="Nori Satoh unpublished cDNA library, tailbud embryo"

ORIGIN

Query Match 3.7%; Score 45; DB 5; Length 711;
Best Local Similarity 47.3%; Pred. No. 0.037;
Matches 169; Conservative 0; Mismatches 185; Indels 3; Gaps 1;
QY 773 ATGTTTGGCTGACCGTTCGTTAGTTCCGAGAGCCATTGCGGAGACATTCGCTTATAAAC 832

DB 206 ATGATCTGTCCAACCTCCCCACCTTACATTATGATCAAGAAAAGTTTGGCCCAATATC 265
QY 833 CGCGGTTCACTGATTCGGGGAGCTGTGCCAAGATACAGTGTGCGCGGTATGGAA 892
DB 266 CACCAATTCCTCAATCAATTTTCGATAAACTTAAACAAAGATATTGAAGTTGATGGAAAGACCA 325
QY 893 TCAAAAAGATACGATTCGTTTTCGTTATGATCGGTGCGCTAACCGGGACCTCGAAGCAT 952
DB 326 TTGTGAAGAACCAATGTTGTTTACATATTTATGATCATCATCATCAGGAGGTTTTT 385
QY 953 TTGAACAGCCTGACGTGTTTAAATATTCATCGGGAAGATCTTGGTATCAAGAGCGCTTTTA 1012
DB 386 GGAAGGATCCTCATATATTTGATCCAGTCGTTTCAGCGCAGGAAAAACATGAAACATCA 445
QY 1013 GCGGCGCGCGCGGATCTCGCTTCGGATCCGGATTCATTAACCTGTGTAGAGAGCTTT 1072
DB 446 ATAGTTATGCTT---ATGTACCTTTCTCTGCTGGCCCAAGAAACTGCATCGCGCAAAAAT 502
QY 1073 TTGCCAAAAGCAATCGAAATTTAGCTAATATTGTGCTGATAGATGCGGAATA 1129
DB 503 TTGCATGAACAAGATGAAATTTGCGGTTCCTCAGGTGTTGAGGCAATTCAGATTA 559

RESULT 3

BE585802/c

LOCUS BE585802

DEFINITION

697 bp mRNA linear EST 17-AUG-2000

cDNA library Triticum aestivum cDNA clone Est#2pT7_B07_b7_057, mRNA

sequence.

ACCESSION BE585802

VERSION BE585802.1 GI:9838834

KEYWORDS EST.

SOURCE

ORGANISM

Triticum aestivum (bread wheat)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Pooideae; Triticeae; Triticum.

1 (bases 1 to 697)

Fellers,J.P., Li,W.L., Hill-Ambroz,K., Matthews,A. and Gill,B.S.

The structure and function of the expressed portion of the wheat

genomes - Kansas State University. Fusarium graminearum infected

spike cDNA library

Unpublished (2000)

Contact: John Fellers

US Department of Agriculture, Agriculture Research Service, Plant

Science and Entomology Unit

Dept. of Plant Pathology, 4006 Throckmorton Hall, Kansas State

University, Manhattan, KS 66506, USA

Tel: 785-532-2367

Fax: 785-532-6167

Email: jpf@alfalfa.ksu.edu

Sequence have been trimmed to remove vector sequence and low

quality sequence with phred score less than 20

Seq primer: 17.

Location/Qualifiers

1. .697

/organism="Triticum aestivum"

/mol_type="mRNA"

/cultivar="Sumai3"

/db_xref="taxon:4565"

/clone="Est#2pT7_B07_b7_057"

/tissue_type="Spike"

/dev_stage="Adult plant"

/lab_host="E. coli JM109"

/clone_lib="KSU wheat Fusarium graminearum infected spike

cDNA library"

/note="Vector: pGEM-T easy; Site_1: SacII; Site_2: SpeI;

Plants were grown in the greenhouse. Spikes were sprayed

with Fusarium graminearum (at what stage). Total RNA, and

poly(A) RNA were prepared from infected spikes. cDNA was

prepared using the SmartTM PCR cDNA synthesis kit from

Clontech. cDNA was cloned into the pGEM-T easy vector

```

ORIGIN
Query Match      3.6%; Score 44.2; DB 2; Length 697;
Best Local Similarity 51.2%; Pred. No. 0.066;
Matches 103; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 775 GTTTGGCTGACGTTGTTAGTTCGAGAGCAATTCGCGAGACATGCGGTATATAACCG 834
Db 617 GATCTTGGGAACTCCACTACTTCTGCGAGTGTATCAAGAGGTTCTTAGATTGCATCCA 558

QY 835 CCGGTTCACTGATTCGCGGCGAGCTGCCAGATACAGTGTCTGCGGCGTATGGAATC 894
Db 557 CTGTTGCGGTTAATCCCTCGAGGACTAGAGAGATTGCAATATATGAGGTTATGACATT 498

QY 895 AAAAAAGATACGATTGTTTTTTTGTATGATCGGTGCGGCTAAACGGGACCTCGAAGCATTT 954
Db 497 CCCAAGATACAAATATATTCATTATACCTTTGCAACATCCCGGATCTCAATATTGG 438

QY 955 GAACAGCTGAGCTGTTTAAT 975
Db 437 GACAGTCTCGAAGTGTTTAAT 417

RESULT 4
CB827487
LOCUS
DEFINITION
LjNEST75d12r Lotus japonicus nodule library 5 and 7 week-old Lotus
corniculatus var. japonicus cDNA 5', mRNA sequence.
ACCESSION
CB827487
VERSION
CB827487.1 GI:29966503
SOURCE
Lotus corniculatus var. japonicus (Lotus japonicus)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae;
Lotus.
REFERENCE
1 (bases 1 to 569)
Colebatch,G., Freund,S., Trevasakis,B and Udvardi,M.
Lotus japonicus root nodule ESTs: tools for functional genomics
Unpublished (2000).
CONTACT: Udvardi MK
Molecular Plant Nutrition
Max Planck Institute of Molecular Plant Physiology
Am Muehlenberg 1, 14476 Golm, Germany
Fax: 49 331 567 8250
Email: udvardi@pimp-golm.mpg.de
Seq primer: T7
High quality sequence/stop: 569.
FEATURES
source
Location/Qualifiers
1..569
/organism="Lotus corniculatus var. japonicus"
/mol_type="mRNA"
/cultivar="Gifu (B-129)"
/db_xref="taxon:34305"
/dev_stage="5 and 7 week-old plants"
/clone_lib="Lotus japonicus nodule library 5 and 7
week-old"
/Note="Organ: Nodule; Vector: pSPORT1; Site 1: Sali;
Site 2: NotI; The library was prepared using mRNA
extracted from nodules of 5 and 7 week-old Lotus plants.
Nodules were induced by, and contained Mesorhizobium
strain R7A."

ORIGIN
Query Match      3.6%; Score 44; DB 6; Length 569;
Best Local Similarity 52.8%; Pred. No. 0.071;
Matches 95; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 804 AGCATTCGGAGACATTCGGTTATAAACGGCGGTTTCAGTATTCCCGGCGGACGTGTC 863
Db 389 AGTTATAAAGAGACACTCAGATTTTATCGCTGGCGCACTAATACCAAGAGAAACAAT 448

from Promega."

QY 864 CCAGATACAGTGGTGGCGGTATGAAATCAAAAAGATACGATGTTTTTTTGTATGAT 923
Db 449 TAAAAATATTATCGTAGATGGGTATGAAATACCAAGCAAAATGTTTATGTAACGT 508

QY 924 CGTGGCGGTACCGGACCCCTGAAGCATTTTGAACAGCCCTGACGTTGTTAAATATTCATCG 983
Db 509 TTGGGCCATCCATAGGATCCCTGAGGCTTGGAAAGACCCACACAGAGTTTAATCTCTGATAG 568

RESULT 5
CNS00396/c
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL063921
VERSION
AL063921.1 GI:4941778
KEYWORDS
GSS.
SOURCE
Drosophila melanogaster (fruit fly)
ORGANISM
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammosser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain v2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source
Location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08K10"
/clone_lib="RPCI-98"
/Note="end : TET3"

ORIGIN
Query Match      3.5%; Score 42; DB 9; Length 1101;
Best Local Similarity 14.6%; Pred. No. 0.37;
Matches 79; Conservative 240; Mismatches 222; Indels 1; Gaps 1;

QY 244 GAAAAGACATCTCTGCCAAAGAGAATTGTAGTGAGAGCTTTATCGTGACGCACTG 303
Db 1101 GKARRWGDDTWRDRTRKDDWMTKWTWTKRDRRWAAGDADRWAWDDGAGTWTATW 1042

QY 304 GATCATCTCTCCATTGATTAAACAAAATCGAAAACTTGTTCGCGCTTATCTTGAA 363
Db 1041 WWWWWATWDTWMDKWWWWATAAKTDTATWWTATRADRWAGRDGRADRDADADG 982

QY 364 AGAGGAAAGTGCATCTCGTCAATGATTTTGGAAAGACGTTTTCGGGTGCGTCACGATG 423
Db 981 AGRDGRKRKDKDRKDDDKGKKKKAKAATKWWDDWDDWDDWDDWDDWDDWDDWDDW 922

QY 424 GCATCTCGGCTCGATATAAAGACACCATGAAAAAATCTCTGAGTGGCACACGGAGTT 483
Db 922
```

```
Db 921 ADDDDGAGKDDGKGKDDDDTDDGTGKDDDDKDDKDDWDKAKGTWGDATWAAATDWWWW 862
Qy 484 GCCGATTTTATCAGAGATATCTCTCAATCTCTGAAGCCGGGCGACATTCGTTATGGTGC 543
Db 861 GWADADWTTWDAADDDWADDDRDWADWAWKDDWAWGARTADRRDWRGRGGRKRR 802
Qy 544 AGCGAACAGCTTCCCAATFACITTGATCGCGGTCAITTAAGAAGCGTCGGCTCAATCCGGGA 603
Db 801 DKRADDDKRAADDDRDADTTWTTTTRDDDDKWKTDWTRWAADRTWDRDDDDDD 742
Qy 604 TCAGATTTAAATTCGATCTCTATGTACTCTGAATGAAGGCATCGCGTGCAGACAAG 663
Db 741 RAGTAGRWRTWKRRKWRDTRWDDADADTDARDDRRRGDDGADAG-KGKKTGRKRR 683
Qy 664 GATATACCTCGCACTGATTTCTTAATGTGCTGTAGCGGCAACGGAACCGCTGATAGAGC 723
Db 682 RDRATWDRTDAAWADAAMWTTTDDTDWDKDRRRKGARRRRRTTARAAMDWMTWKAWD 623
Qy 724 CTGGCACTGATCATCTACCATTTGCTCAACAATCTCGAGCAGATGAATGATGTTTGGCT 783
Db 622 WAKWDKTRADWRDWAADTWTDDKADRDWAKARAWRRDRARARADRRWTTGKTT 563
Qy 784 GA 785
Db 562 TA 561

RESULT 6
AL823229 439 bp mRNA linear EST 15-JUL-2002
LOCUS AL823229 p:133 Triticum aestivum cDNA clone C04_p133_plate_1, mRNA
DEFINITION sequence.
ACCESSION AL823229
VERSION AL823229.1 GI:21834998
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 439)
AUTHORS Wilson,I., Beswick,R., Shepherd,S., Barker,G., Parker,J., Owen,P.,
Edwards,D., Coghill,J., Holdsworth,M., Lenton,J., Shewry,P. and
Edwards,K.
TITLE A BBSRC-funded wheat EST resource for the academic community
JOURNAL Unpublished (2002)
COMMENT Contact: Barker G
Institute of Arable Crop Research
Long Ashton, Bristol BS41 9AF United Kingdom.
FEATURES
source
1..439
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Mercia"
/db_xref="taxon:4565"
/clone="C04_p133_plate_1"
/tissue_type="cold stressed seedlings"
/dev_stage="21 days old"
/clone_lib="p:133"

ORIGIN
Query Match 3.4%; Score 41.8; DB 1; Length 439;
Best Local Similarity 50.2%; Pred. No. 0.32;
Matches 103; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

Qy 769 AATGATGTTTTCGGCTGACCGTTCGTTAGTTCGAGAGCCATTGCGGAGACATTCGGTTAT 828
Db 168 AATAAGCACTCGTGAATCGACTACACGAGATAGTATCAAGAAACCTCAGGTTTC 227
Qy 829 AAACCGCGGTTTCACTGATTCGCGGCACTGTCTCCAAAGATACATGCTCGCGGTATG 888
Db 228 CATCGCCTGCTCTCTGATTCCTCGCGCTCCATAGAGGATTGTACTATCATGGTTAT 287
```

```
Qy 889 GAAATCAAAAAGATACGATTGTTTTTTTGTATGATCGGTGCGCTAAACCGGACCCCTGAA 948
Db 288 GACATCCCCCAAGGACCACCAATGATATACGTTAATGTTCTTCGCGATTTCTCGAGATCTGCA 347
Qy 949 GCATTTGAACAGCCCTGACGTGTTTA 973
Db 348 TACTGTAAACAATCTCTTGGGAGTTTA 372

RESULT 7
AL174775 330 bp DNA linear GSS 13-JAN-2003
LOCUS AL174775
DEFINITION Mycobacterium avium subsp. paratuberculosis DNA
Mycobacterium avium subsp. paratuberculosis genomic clone
Mptb61.15, genomic survey sequence.
ACCESSION AL174775
VERSION AL174775
KEYWORDS GSS.
SOURCE Mycobacterium avium subsp. paratuberculosis
ORGANISM Mycobacterium avium subsp. paratuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
avium complex (MAC).
REFERENCE 1 (bases 1 to 330)
AUTHORS Nielsen,K.K. and Ahrens,P.
TITLE Subtractive hybridization and bioinformatics: combining genetic
methods to obtain new information on species-specific elements of
Mycobacterium avium subspecies paratuberculosis
JOURNAL Unpublished (2003)
COMMENT Contact: Nielsen KK
Bacteriology
Danish Veterinary Institute
Bulowsevej 27, Copenhagen V, 1790, Denmark
Email: kni@vdi.dk
Class: unknown.
FEATURES
source
1..330
/organism="Mycobacterium avium subsp. paratuberculosis"
/mol_type="genomic DNA"
/sub_species="paratuberculosis"
/db_xref="taxon:1770"
/clone="Mptb61.15"
/clone_lib="Mycobacterium avium subsp. paratuberculosis
DNA"

ORIGIN
Query Match 3.4%; Score 41.6; DB 8; Length 330;
Best Local Similarity 46.3%; Pred. No. 0.33;
Matches 137; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

Qy 589 CGCGTCAATCCGGGATCAGATTAAATTTTCGATCTCTATGTTCTTGAATATGAAGCATG 648
Db 29 CGCGCGGAGCCAGGTGACGATATCGCGACTTCTTACTCCACGCGAGGTTGATGGTCAG 88
Qy 649 GCGCTGTGCGACAAGATATCTCGCACTGATTCTTAATGTCGTGTTAGCCGCAACGGAA 708
Db 89 AGCCTCAGGATTGGAAATTTAACTGTTCTTTATGTTGCTGATCGATCCCTGAGTGAT 148
Qy 709 CCGGCTGATAGACGCTGGCACTGATGATCTACCAATTTGCTCAACAATCTCTGAGCAGATG 768
Db 149 ACGACCCCGCAATCTTTGTAGCAGCGGGAATTCCTGCTGCTGCTGAGCATCCCAAGAACTG 208
Qy 769 AATGATGTTTGGCTGACGTTTCGTTAGTTTCGAGAGCAATTCGCGAGACATTCGCTTAT 828
Db 209 CAGCGGCTTAAGGACAGACCTTCACTCATGCGGACCGGATCGAAGAAATGTCGCTAC 268
Qy 829 AAACCGCGGTTCACTGATTCGCGGCGAGCTGCCAAGATACAGTGTGTCGCGCG 884
Db 269 ACGTCGCGGTGACGGCATTTCTGCGCACCGCGACCAAGACACTGAGTTGCGCG 324

RESULT 8
CK535040
```


LOCUS CK535040 631 bp mRNA linear EST 15-JAN-2004
DEFINITION rswd00_003392.y1 swg Bombyx mori cDNA, mRNA sequence.
ACCESSION CK535040
VERSION CK535040.1 GI:40919494
KEYWORDS Bombyx mori (domestic silkworm)
SOURCE Bombyx mori
ORGANISM Bombyx mori

REFERENCE 1 (bases 1 to 631)
AUTHORS Xia,Q., Zhou,Z., Lu,C., Cheng,D., Dai,F., Li,B., Zhao,P., Zha,X.,
Cheng,T., Chai,C., Pan,G., Xu,J., Liu,C., Lin,Y., Qian,J., Hou,Y.,
Wu,Z., Li,G., Pan,M., Li,C., Shen,Y., Lan,X., Yuan,L., Yang,G.,
Wan,Y., Zhu,Y., Yu,M., Shen,W., Wu,D., Xiang,Z., Yu,J., Wang,J.,
Li,R., Li,H., Li,G., Su,J., Wang,X., Li,G., Zhang,Z., Li,J.,
Zhang,Q., Wei,N., Xu,J., Sun,H., Dong,L., Liu,D., Zhao,S., Zhao,X.,
Meng,Q., Lan,F., Fang,L., Li,Y., Wang,J., Ye,J., Wong,G.K.-S.,
Yang,H., Shi,J., Wu,Q., Li,C., Li,D., Sun,Y., Zhang,Z., Yang,Z.,
Huang,Y., Xi,Y., Qi,Q., He,D., Huang,H., Zhang,X., Wang,Z., Li,W.
and Cao,Y.

TITLE A draft sequence for the genome of the domesticated silkworm
(Bombyx mori)

JOURNAL Unpublished (2004)
COMMENT Contact: Yang Huan Ming
Beijing Genomics Institute
Chinese Academy of Sciences
Beijing Airport Industrial Zone B6, Beijing 101300, P.R.China
Email: hyang@igtp.ac.cn.

FEATURES
source
1..631
/organism="Bombyx mori"
/mol_type="mRNA"
/strain="DaZhao(P50)"
/db_xref="taxon:7091"
/sex="male"
/tissue_type="testis"
/dev_stage="5th-instar day-3 larva"
/clone_lib="swg"
/note="Vector: pBluescript II SK(+)"

ORIGIN
Query Match 3.3%; Score 40.6; DB 7; Length 631;
Best Local Similarity 48.9%; Pred. No. 0.85;
Matches 109; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 768 GAATGATGTTTGGCTGACCGTTCTGTTACTTCGAGAGCCATTGCGGAGACATTGGTTA 827
Db 189 GTACAAATGAACGGGACAAATGAAGTTCTGAGTGGGTCAACGAATCGTTGCGGAT 248

QY 828 TAAACCGCCGGTTACGCTATTCGCGGCGAGCTGTCCCAAGATACAGTGGTGGCGGGTAT 887
Db 249 GTATCCATCGGTACCGCTGATCGAGTTGGATCACAAGAGTGTGAGGTGGAGTCT 308

QY 888 GGAATCAAAAAGATACGATTGTTTTTGTATGATCGTGGCGGTAAACCGGAGCCCTGA 947
Db 309 GAAACTATCTAAAGGAACATCGTCCGGTTAAACATATTCCAGATGCATCGAAATCCAGA 368

QY 948 AGCATTTGAACAGCCCTGAGCTGTTTATATTATTCATCGGGAAGAT 990
Db 369 AGGGTTGAGNAGCCCTCGAATTCTTCTCTGAACGATTCGAT 411

RESULT 9
CA923083/c
LOCUS CA923083 MTUS Medicago truncatula cDNA clone MTUS-62F12, mRNA
DEFINITION EST640801
VERSION CA923083
ACCESSION CA923083.1 GI:27410013
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula

LOCUS CF347163 747 bp mRNA linear EST 18-AUG-2003
DEFINITION AGENCOURT 15223896 NIH_ZGC_10 Danio rerio cDNA clone IMAGE:7002517
5', mRNA sequence.
ACCESSION CF347163
VERSION CF347163.1 GI:33789768
KEYWORDS EST.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 747)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.

REFERENCE 1 (bases 1 to 777)
AUTHORS VandenBosch,K., Endre,G., Silverstein,K., Town,C.D., Van Aken,S.,
Utterback,T., Cheung,F. and Fraser,C.M.
TITLE The Medicago truncatula 6K unigene set: cDNA clones selected and
re-arrayed from various libraries
JOURNAL Unpublished (2002)
COMMENT Contact: VandenBosch K
Department of Plant Biology
University of Minnesota
220 Biosci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
Tel: 612 624 2755
Fax: 612 625 1738
Email: kvandenb@cbs.umn.edu
Alias Clone NF013H10IN
TIGR sequence name: MTUCJ72TV
More information is available at: www.medicago.org
Seq primer: (gtA ATA CGA Ctc Act ATA ggg C).

FEATURES
source
1..777
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="A17"
/db_xref="taxon:3880"
/clone_lib="MTUS-62F12"
/tissue_type="mixed tissues"
/dev_stage="various stages"
/lab_host="XLOLR"
/clone_lib="MTUS"
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the Unizap XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-Zap phage using Ex-assist
helper phage and propagated in XLOLR cells."

ORIGIN
Query Match 3.3%; Score 40.6; DB 6; Length 777;
Best Local Similarity 52.7%; Pred. No. 0.91;
Matches 88; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 807 CATTGCGGAGACATTGCGTTATAAACCGCGCTTCAGCTGATTCGCGGCGAGTGTCCCA 866
Db 471 CATAAAGAACACATGAGGCTTCATCCCGATACCGATATATGAGGAAGGAATGGA 412

QY 867 AGATACAGTGTGCGCGGTATGGAATCAAAAAGATACGATGTTTTTTTGTATGATCGG 926
Db 411 AGATTGTGTGGTTGATGGTTAAACATGATTCGAAAGGCTCAATGGTTGTGTAACATTG 352

QY 927 TCGCGCTAACCGGACCCCTGAAGCAATTTGAACAGCTGACGTGTTTA 973
Db 351 GGCTATGGCAAGGAGCCCAAGATATGGGAAACCCCATTTAGAGTTTA 305

RESULT 10
CF347163
LOCUS CF347163 747 bp mRNA linear EST 18-AUG-2003
DEFINITION AGENCOURT 15223896 NIH_ZGC_10 Danio rerio cDNA clone IMAGE:7002517
5', mRNA sequence.
ACCESSION CF347163
VERSION CF347163.1 GI:33789768
KEYWORDS EST.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 747)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

Query Match	3.2%;	Score 39.4;	DB 5;	Length 779;
Best Local Similarity	62.9%;	Pred. No. 2.2;	36;	Indels 0;
Matches	61;	Conservative	0;	Mismatches
QY	1064	GAGCAGCTTTTGGCCAAAAACGAAATCGAAATTGTAGCTAATATTGTGCTGGATAAGATGC	1123	
DB	348	GAGAAAAATTCGCGAAAAACGAAACGACATTATATAGTATTGTGCTGTAAAACTGGC	407	
QY	1124	GGAATATCATGATTAGAGGAGATTTTTTGTATGCTGA	1160	
DB	408	TAAAGAACAGTTTACAAAATGATGATGATGCGGA	444	
RESULT 14				
BE432743				
LOCUS	BE432743			
DEFINITION	EST392972 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA clone cLEG1019, mRNA sequence.			
ACCESSION	BE432743			
VERSION	BE432743.1	GI:9430586		
KEYWORDS	EST.			
SOURCE	Lycopersicon esculentum (tomato)			
ORGANISM	Lycopersicon esculentum			
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.			
AUTHORS	1 (bases 1 to 585) Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E., Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Rønning,C.M., Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.			
TITLE	Generation of ESTs from tomato fruit tissue, breaker stage			
JOURNAL	Unpublished (2000)			
COMMENT	Contact: CUGI Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Email: http://www.genome.clemson.edu/orders/index.html 5 prime sequence.			
FEATURES				
source	Location/Qualifiers 1..585 /organism="Lycopersicon esculentum" /mol_type="mRNA" /cultivar="TA496" /db_xref="taxon:4081" /clone="cLEG1019" /tissue_type="Pericarp" /dev_stage="breaker" /lab_host="SOLR" /clone_lib="tomato breaker fruit, TIGR" /note="Vector: pBluescriptSKmCuadapt; Site 1: EcoR1; Site 2: XhoI; Fruit were harvested at the breaker stage (first sign of lycopene accumulation on the blossom end of the fruit). Fruit were cut in half and the seeds and locules were discarded prior to freezing the pericarp."			
ORIGIN				
Query Match	3.2%;	Score 39.2;	DB 2;	Length 585;
Best Local Similarity	49.1%;	Pred. No. 2.3;		
Matches	104;	Conservative	0;	Mismatches
QY	804	AGCCATTGCGGAGACATTGCGTTATAAACCGCGGTTTCAGCTCATTCGCGGAGCTGTC	863	
DB	10	AGTTATAAAGAGACATTGAAATTGTATCCACCGCTCCATTATTAACGAGAGACAAT	69	
QY	864	CCAAGATACAGTGGTCGCGCGGTATGGAAATCAAAAAAGATACGATTGTTTTTGTATGAT	923	
DB	70	GGAAAAATGCATCATAGATGGATGATGAATCAAAACAGATACTCCAGTTTCTATCAACGC	129	
QY	924	CGGTGCGGCTAACCGGGACCGCTGAAGCATTGGAACAGCCTGACGTGTTTAATTATTCATCG	983	
DB	130	GTGGGCAATAGGTAGAGATCCTGAATGTTGGGAAAAATCCAGACGAGTTTATTCCTGAGAG	189	

source		1. .930		/organism="Danio rerio"		/mol_type="mRNA"		/db_xref="taxon:7955"		/clone="IMAGE:6997518"		/tissue_type="whole body"		/lab_host="DH10B"		/clone_lib="NIH ZGC 7"		/note="Vector: pExpress1; Site 1: NotI; Site 2: EcoRV; Bulk tissue was collected from a whole adult individual from the Tuebingen strain. 1st strand cDNA was primed with a Not I - oligo(dT) primer, double-stranded cDNA was cloned into the Not I and EcoRV sites of pExpress-1. Library was size-selected for >1 kb fragments and normalized. A non-normalized version of this library is also available (NIH_ZGC10). Library was constructed by Open Biosystems (Huntsville, AL)"																																																							
ORIGIN		Query Match		3.3%;		Score 40; DB 6; Length 930;		Best Local Similarity		58.3%;		Pred. No. 1.5; Mismatches 0; Indels 0; Gaps 0;		Matches 70; Conservative																																																											
QY		547		GAACAGCTTTCCCACTACTGATCGCGGTCACTTAAAGAACGTCGCGTCAATCCGGGATCA		606		Db		345		GAACCTTCATTACCAAACTTGTTGCGGCTGATTAATGTTTTCAGCGGCGCATTTTA		404		QY		607		GATTTAATTCGATCCTATGCTACTTCTGAATATGAGGCATGGCGCTGTCGGACAAAGAT		666		Db		405		GGCCTAGTAGTGACACATGCTGTATAGATAATAAAGTCATGGTGCAATGGTGGAAAAAGAAAT		464																																											
RESULT 13		BU796025		LOCUS		SUF2DXG07 SJF Schistosoma japonicum cDNA similar to pir JC1196		DEFINITION		protein phosphatase inhibitor 2 - rabbit, mRNA sequence.		ACCESSION		BU796025.1		GI:29353032		EST.		Schistosoma japonicum		ORGANISM		Schistosoma japonicum		REFERENCE		Hu, W., Fan, Q., Shen, D. K., Liu, F., Zhu, Z. D., Song, H. D., Xu, X. R., Wang, Z. J., Rong, Y. P., Zeng, L. C., Wu, J., Zhang, X., Wang, J. J., Xu, X. N., Wang, S. Y., Fu, G., Zhang, X. L., Wang, Z. Q., Brindley, P. J., McManus, D. P., Xue, C. L., Feng, Z., Chen, Z. and Han, Z. G.		Evolutionary and biomedical implications of a Schistosoma japonicum complementary DNA resource		Nat. Genet. 35 (2), 139-147 (2003)		22879925		12973349		COMMENT		Contact: Zequang Han		Chinese National Human Genome Center at Shanghai		351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai		201203, P. R. China		Tel: 86-21-50801919 (ex.45)		Fax: 86-21-50801922		Email: hanzg@chgc.sh.cn.		Location/Qualifiers		1. .779		/organism="Schistosoma japonicum"		/mol_type="mRNA"		/db_xref="taxon:6182"		/sex="female"		/tissue_type="whole body"		/dev_stage="adult"		/lab_host="rabbits"		/clone_lib="SJF"	
FEATURES		source		Query Match		3.2%;		Score 39.2; DB 2; Length 585;		Best Local Similarity		49.1%;		Pred. No. 2.3; Mismatches 104; Conservative		0; Indels 0; Gaps 0;		Matches 104; Conservative				QY		804		ACCCATTGCGGAGACATTGCGTTATTAACCCGCGTTCAGCTGATTCGGCGGACGTGTC		863		Db		10		ACTTATAAGAGACATTGAAATTTGATCCACCGCTCCATTATACAGAGAGACAAT		69		QY		864		CCAAGATACAGTGGTCGGCGGTATGGAAATCAAAAAAGATACGATTTGTTTTGTATGAT		923		Db		70		GGAAAAATGTCATCATAGATGGATGGAATCAACAGATACTCCAGTTCTATCAACGC		129		QY		924		CGTGGCGGTACACCGGACCTCGAAGCATTTTGAACAGCTGACGCTGTTTATTCATCG		983		Db		130		GTGGGCAATAGGTAGAGATCCTGATGTGGGAAATCCAGACGAGTTATTTCTGAGAG		189					
ORIGIN		Query Match		3.2%;		Score 39.4; DB 5; Length 779;		Best Local Similarity		62.9%;		Pred. No. 2.2; Mismatches 61; Conservative		0; Indels 0; Gaps 0;		Matches 61; Conservative						QY		1064		GAGCAGCTTTTCCCAAAACGAAATCGAAATGTAGCTAATATTGTCTCGATAAGATGC		1123		Db		348		GAGAAATTTCCGCGAAAAACGAAAGCAACATTATAATGAGTATTTTGCTCTGTAACCTGGC		407		QY		1124		GGAATATCAGATTAGAGGAACATTTTGTATTGCTGA		1160		Db		408		TAAAGACAGTTTACAAATGATGATGATGCGGA		444																					
RESULT 14		BE432743		LOCUS		EST199272 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA clone cLEG1019, mRNA sequence.		DEFINITION		clone cLEG1019, mRNA sequence.		ACCESSION		BE432743		GI:9430586		VERSION		Lycopersicon esculentum (tomato)		ORGANISM		Lycopersicon esculentum		REFERENCE		Alcala, J., Vrebalov, J., White, R., van der Hoeven, R. S., Holt, I. E., Liang, F., Hansen, T. S., Craven, M. B., Bowman, C. L., Rønning, C. M., Nierman, W., Fraser, C. M., Martin, G. B., Giovannoni, J. J. and Tanksley, S. D.		Generation of ESTs from tomato fruit tissue, breaker stage		Unpublished (2000)		Contact: CUGI		Clemson University Genomics Institute		Clemson University		100 Jordan Hall, Clemson, SC 29634, USA		Email: http://www.genome.clemson.edu/orders/index.html		5 prime sequence.		Location/Qualifiers		1. .585		/organism="Lycopersicon esculentum"		/mol_type="mRNA"		/cultivar="TA496"		/db_xref="taxon:4081"		/clone="cLEG1019"		/tissue_type="pericarp"		/dev_stage="breaker"		/lab_host="SOLR"		/clone_lib="tomato breaker fruit, TIGR"		/note="Vector: pBluescriptSKMuadapt; Site 1: EcoRI; Site 2: XhoI; Fruit were harvested at the breaker stage of (first sign of lycopene accumulation on the blossom end of the fruit). Fruit were cut in half and the seeds and locules were discarded prior to freezing the pericarp."					

GenCore version 5.1.6
Copyright (c) 1993 - 2004. Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2004, 13:19:06 ; Search time 122.858 Seconds
(without alignments)
7029.309 Million cell updates/sec

Title: US-10-627-124-1
Perfect score: 1215
Sequence: 1 atggccaatcgattaaatt.....tcgtgcttgtagcgggsgca 1215

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	67.8	5.6	4403765	3	US-09-103-840A-2
C 2	67.8	5.6	4411529	3	US-09-103-840A-1
C 3	53.6	4.4	4403765	3	US-09-103-840A-2
C 4	53.6	4.4	4411529	3	US-09-103-840A-1
C 5	50	4.1	47981	4	US-09-679-279-1
C 6	49.6	4.1	6085	3	US-09-029-603-4
C 7	47.6	3.9	1209	4	US-09-724-797-35
C 8	44.6	3.7	8478	3	US-08-676-818-1
C 9	44.6	3.7	8478	3	US-09-407-549-1
C 10	43	3.5	1209	3	US-09-105-537-21
C 11	43	3.5	13613	3	US-09-105-537-3
C 12	43	3.5	38506	3	US-09-320-878-19
C 13	43	3.5	38506	4	US-09-141-908-1
C 14	43	3.5	38506	4	US-09-657-440-19
C 15	42.4	3.5	37556	2	US-08-576-626A-1
C 16	41	3.4	37030	4	US-08-311-731A-25
C 17	40.6	3.3	1251	3	US-09-105-537-38
C 18	40.6	3.3	5970	3	US-09-320-878-21
C 19	40.6	3.3	5970	4	US-09-141-908-11
C 20	40.6	3.3	5970	4	US-09-657-440-21
C 21	37.8	3.1	1212	6	5212296-17
C 22	37.4	3.1	1735	5	US-08-102-863-10
C 23	37.4	3.1	1735	5	PCT-US92-10885-10
C 24	37	3.0	77536	4	US-09-410-551B-1
C 25	37	3.0	77536	4	US-09-940-316B-1
C 26	35	3.0	536165	4	US-09-214-808-1
C 27	35	2.9	4084	2	US-08-568-459A-1

ALIGNMENTS

RESULT 1

US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-08-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match	5.6%	Score 67.8;	DB 3;	Length 4403765;
Best Local Similarity	45.1%	Pred. No. 1.6e-09;		
Mismatches	292;	Conservative	0;	Mismatches 352;
Indels	3;	Gaps	1;	
QY	337	GAAACTTTGTTAGCCCTTATCTTGAAGAGGAAAGTATCTCGTCAATGATTTTGA	396	
DB	3947917	GACACCTTGATCGACGCCGTGTGCGAAGCGCGAGTGTGATCTCGTGGGACCTGGCC	3947858	
QY	397	AAGAGCTTTCGGTGTGGTCAAGATGACATGTCGGGCTGGATAAAGAGACCATGAA	456	
DB	3947857	GGCGCGCTACCGATGGCGGTGATCGGCACATGCTCGGGGTCCGTCAGAGCGGGAC	3947798	
QY	457	AAATCTCTGAGTGGCAGACGGAGTTCGGGATTTTATCAGAGTATCTCTCAATCTCT	516	
DB	3947797	ATGTTCTTTCGGTGTGGTCCGACGATCTGGTGACATCTCTCCTCCTCA	3947741	
QY	517	GAAGCGCGGCACATTCGTTATGTTGTCGAGCAAGCTTTCCCAATCTGATGCGGTC	576	
DB	3947740	GAGGATTTCCGATCACCATGACGACCTTCGCGGCTTACACGACTTACCCCGGCCAC	3947681	
QY	577	ATTAAGAACGTCGCGTCAATCCGGGATCAGATTTAAATTTTCGATCCTATGTTACTT	636	
DB	3947680	ATTGCGGACGCGGAGCGGACCCACCGACCTGGTGGTGGTGGTGGTGGTGGTGGT	3947621	

QY 948 AGCATTTGAACAGCTGACGTTGTTTAAATATTCATCGGAA 987
Db 874566 CGTGTTCGACGGCGGAGTTCGATATCACCCGAAA 874605

RESULT 4
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 4.4%; Score 53.6; DB 3; Length 4411529;
Best Local Similarity 47.4%; Pred. No. 7.3e-05;
Matches 161; Conservative 0; Mismatches 179; Indels 0; Gaps 0;
QY 648 GCGCGTGTGCGACAGGATATACCTGCACTGATTTCTTAATGTCTTAGCGCGCAACGGA 707
Db 872136 GCGCGTGTGCGACGCGCACTGACCTGTTCTTCACTACTGTTTCAGCGCGCGCGGA 872195
QY 708 ACCGGCTGATAGACGCTGCGACTGATGATACCAATTTGCTCAACAATCTTGGAGCAGAT 767
Db 872196 AACCACCGTAACTCCATTCGCGGGGCTGCTGGCGTGGCGAGAACCTTGACCACT 872255
QY 768 GAATGATGTTGGTGACGCTGCTAGTTCGAGAGCATTGCGGAGACATCTGGGTTA 827
Db 872256 GCAACGCTGCGAAGCGATTTTGAAGTGTTCGCGACTGCGATCGAAGATCGTGAGGTG 872315
QY 828 TAAACGCGGGTTACGCTGATTCGCGGCGAGCTGCCCAAGATACAGTGTGCGGCGTAT 887
Db 872316 GAGTCGCGCTCACATCGAAGCGGCGACGGGCTCCGTCGCTGAGCTGCGCGCGCA 872375
QY 888 GGAAATCAAAAAGATACGATTTGTTTGTATGATCGGTGCGGCTAAACGGGACCCCTGA 947
Db 872376 GCGATCGAGGCGGGTCAGAAGTGTGTTGGAGGGCTCGGCCAACCGTGATCCAG 872435
QY 948 AGCATTTGAACAGCTGACGTTGTTTAAATATTCATCGGAA 987
Db 872436 CGTGTTCGACGGCGGAGTTCGATATCACCCGAAA 872475

RESULT 5
US-09-679-279-1/c
; Sequence 1, Application US/09679279
; Patent No. 6524841
; GENERAL INFORMATION:
; APPLICANT: McDaniel, Robert
; APPLICANT: Volchegursky, Yanina
; TITLE OF INVENTION: Recombinant Megalomicin Biosynthetic
; TITLE OF INVENTION: Genes and Uses Thereof
; FILE REFERENCE: 300622004700
; CURRENT APPLICATION NUMBER: US/09/679,279
; CURRENT FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/158,305
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: US 60/190,024

; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 47981
; TYPE: DNA
; ORGANISM: Micromonospora megalomicea
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(144)
; OTHER INFORMATION: megBVI (megT), TDP-4-keto-6-deoxyglucose-2,3-dehydratase;
; OTHER INFORMATION: SEQ ID NO: 2= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (928)...(2061)
; OTHER INFORMATION: megDVI, TDP-4-keto-6-deoxyglucose 3,4-isomerase,
; OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 3,4-isomerase;
; OTHER INFORMATION: SEQ ID NO: 3= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (2072)...(3382)
; OTHER INFORMATION: megDI, rhodosaminyl transferase (eryCIII homolog),
; OTHER INFORMATION: TDP-megosamine glycosyltransferase;
; OTHER INFORMATION: SEQ ID NO: 4= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (3462)...(4634)
; OTHER INFORMATION: megG (megY), mycarosyl acyltransferase, mycarose O-acyltransferase,
; OTHER INFORMATION: SEQ ID NO: 5= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (4651)...(5775)
; OTHER INFORMATION: megDII, deoxysugar transaminase (eryCI, DnrJ homolog),
; OTHER INFORMATION: TDP-3-keto-6-deoxyhexose 3-aminotransaminase;
; OTHER INFORMATION: SEQ ID NO: 6= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (5822)...(6595)
; OTHER INFORMATION: megDIII, daunosaminyl-N,N-dimethyltransferase (eryCVI homolog);
; OTHER INFORMATION: SEQ ID NO: 7= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (6592)...(7197)
; OTHER INFORMATION: megDIV, TDP-4-keto-6-deoxyglucose 3,5-epimerase (eryBVII, dnmU hom
; OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 3,5-epimerase;
; OTHER INFORMATION: SEQ ID NO: 8= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (7220)...(8206)
; OTHER INFORMATION: megDV, TDP-hexose 4-ketoreductase (eryBIV, dnmV homolog),
; OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;
; OTHER INFORMATION: SEQ ID NO: 9= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (8228)...(9220)
; OTHER INFORMATION: megBII-1 (megDVII), TDP-4-keto-L-6-deoxy-hexose 2,3-reductase;
; OTHER INFORMATION: SEQ ID NO: 10= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (9226)...(10479)
; OTHER INFORMATION: megBV, mycarosyl transferase, mycarose glycosyltransferase;
; OTHER INFORMATION: SEQ ID NO: 11= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (10483)...(11424)
; OTHER INFORMATION: megBIV, TDP-hexose 4-ketoreductase,
; OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;
; OTHER INFORMATION: SEQ ID NO: 12= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (12181)...(22821)
; OTHER INFORMATION: megAI; SEQ ID NO: 13= translated amino acid sequence
; NAME/KEY: misc feature
; LOCATION: (12505)...(13470)
; OTHER INFORMATION: megAI, AT-L
; NAME/KEY: misc feature
; LOCATION: (13576)...(13791)
; OTHER INFORMATION: megAI, ACP-L
; NAME/KEY: misc feature
; LOCATION: (13849)...(15126)
; OTHER INFORMATION: megAI, KSI
; NAME/KEY: misc feature
; LOCATION: (15427)...(16476)
; OTHER INFORMATION: megAI, ATI

; NAME/KEY: misc feature
; LOCATION: (17155)...(17694)
; OTHER INFORMATION: megAI, KR1
; NAME/KEY: misc feature
; LOCATION: (17947)...(18207)
; OTHER INFORMATION: megAI, ACP1
; NAME/KEY: misc feature
; LOCATION: (18258)...(19548)
; OTHER INFORMATION: megAI, KS2
; NAME/KEY: misc feature
; LOCATION: (19876)...(20910)
; OTHER INFORMATION: megAI, AT2
; NAME/KEY: misc feature
; LOCATION: (21517)...(22053)
; OTHER INFORMATION: megAI, KR2
; NAME/KEY: misc feature
; LOCATION: (22318)...(22575)
; OTHER INFORMATION: megAI, ACP2
; NAME/KEY: CDS
; LOCATION: (22867)...(33555)
; OTHER INFORMATION: megAI; SEQ ID NO: 14= translated amino acid sequence
; NAME/KEY: misc feature
; LOCATION: (22957)...(24237)
; OTHER INFORMATION: megAI, KS3
; NAME/KEY: misc feature
; LOCATION: (24544)...(25581)
; OTHER INFORMATION: megAI, AT3
; NAME/KEY: misc feature
; LOCATION: (26230)...(26733)
; OTHER INFORMATION: megAI, KR3 (inactive)
; NAME/KEY: misc feature
; LOCATION: (26938)...(27258)
; OTHER INFORMATION: megAI, ACP3
; NAME/KEY: misc feature
; LOCATION: (27393)...(28590)
; OTHER INFORMATION: megAI, KS4
; NAME/KEY: misc feature
; LOCATION: (28897)...(29931)
; OTHER INFORMATION: megAI, AT4
; NAME/KEY: misc feature
; LOCATION: (29953)...(30477)
; OTHER INFORMATION: megAI, DH4
; NAME/KEY: misc feature
; LOCATION: (31356)...(32244)
; OTHER INFORMATION: megAI, ER4
; NAME/KEY: misc feature
; LOCATION: (32257)...(32799)
; OTHER INFORMATION: megAI, KR4
; NAME/KEY: misc feature
; LOCATION: (33052)...(33312)
; OTHER INFORMATION: megAI, ACP4
; NAME/KEY: CDS
; LOCATION: (33666)...(43271)
; OTHER INFORMATION: megAI; SEQ ID NO: 15= translated amino acid sequence
; NAME/KEY: misc feature
; LOCATION: (33780)...(35027)
; OTHER INFORMATION: megAI, KS5
; NAME/KEY: misc feature
; LOCATION: (35385)...(36419)
; OTHER INFORMATION: megAI, AT5
; NAME/KEY: misc feature
; LOCATION: (37068)...(37604)
; OTHER INFORMATION: megAI, KR5
; NAME/KEY: misc feature
; LOCATION: (37860)...(38120)
; OTHER INFORMATION: megAI, ACP5
; NAME/KEY: misc feature
; LOCATION: (38187)...(39470)
; OTHER INFORMATION: megAI, KS6
; NAME/KEY: misc feature
; LOCATION: (39795)...(40811)
; OTHER INFORMATION: megAI, AT6
; NAME/KEY: misc feature

; LOCATION: (41406)...(41936)
; OTHER INFORMATION: megAI, KR6
; NAME/KEY: misc feature
; LOCATION: (42168)...(42425)
; OTHER INFORMATION: megAI, ACP6
; NAME/KEY: misc feature
; LOCATION: (42585)...(43271)
; OTHER INFORMATION: megAI, TE
; NAME/KEY: CDS
; LOCATION: (43268)...(44344)
; OTHER INFORMATION: megAI, TDP-4-keto-6-deoxyglucose 3,4-isomerase;
; OTHER INFORMATION: SEQ ID NO: 16= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (44355)...(45623)
; OTHER INFORMATION: megAI, desosaminyl transferase, desosamine glycosyltransferase;
; OTHER INFORMATION: SEQ ID NO: 17= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (45620)...(46591)
; OTHER INFORMATION: megAI, TDP-4-keto-6-deoxy-L-glucose 2,3 dehydratase,
; OTHER INFORMATION: TDP-4-keto-6-deoxyglucose 2,3 dehydratase;
; OTHER INFORMATION: SEQ ID NO: 18= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (46660)...(47403)
; OTHER INFORMATION: megAI, TEII; SEQ ID NO: 19= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (47411)...(47980)
; OTHER INFORMATION: megAI, C-6 hydroxylase; SEQ ID NO: 20= translated amino acid sequer
US-09-679-279-1

Query Match 4.1%; Score 50; DB 4; Length 47981;
Best Local Similarity 48.0%; Pred. No. 7.4e-05;
Matches 143; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

QY 689 TGCTTTAGCCGACGAAACCGGCTGATAGACGCTGGCACTGATGATCTACCATTTGC 748
|||||
Db 47913 TGCTGCTGCGGGATTTCGAGGCGTCCGTGAGCCTCATCGGGATCGGCACCTACCTGCTGC 47854
|||||
QY 749 TCAACAATCCTGAGCAGATGATGATTTTGGCTGACCGTTGCTAGTTCCGAGAGCCA 808
|||||
Db 47853 TGACCCACCCGACCACTCGCGCTGGTCCGGAAGACCCCGCACTGCTGCCGAGCGG 47794
|||||
QY 809 TTGCGGAGACATTCGTTTATAAACCGCGGTTTCACTGATTCGCGGAGCTGTCCCAAG 868
|||||
Db 47793 TGGAGGAGATCTGCGCTACCAGGCCCCACCGGAGACACACCCGGTTCCGCCACCGCG 47734
|||||
QY 869 ATACAGTGGTGGCGGATGGAATCAAAAAGATACGATTTTGTATGATCGGTG 928
|||||
Db 47733 AGGTGGAGATCGCGGGGTGACGATCCGCGGCGTACAGCACCGCTCTGTATCGCAACGGG 47674
|||||
QY 929 CGGCTAACCGGACCCCTGAAGCATTGAAACGCTGACGTGTTTAAATTCATCGGA 986
|||||
Db 47673 CGGCCAACCGGACCCCGGACAGTTTCCCGACCCCGACCGGTTCCGACGTCACCAAGGA 47616
|||||

RESULT 6

US-09-029-603-4
; Sequence 4, Application US/09029603
; Patent No. 6210935
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Engel, Natalie
; APPLICANT: Bietenhader, Jurg
; APPLICANT: Roupel, Christine
; APPLICANT: Pospiech, Andreas
; TITLE OF INVENTION: Staurosporin Biosynthesis Gene Clusters
; FILE REFERENCE: 4-20555/A/PCT
; CURRENT APPLICATION NUMBER: US/09/029,603
; CURRENT FILING DATE: 1998-03-20
; EARLIER APPLICATION NUMBER: PCT/EP96/03643
; EARLIER FILING DATE: 1996-08-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4


```
; LENGTH: 6085
; TYPE: DNA
; ORGANISM: Streptomyces longisporoflavus
; FEATURE:
; NAME/KEY: misc RNA
; LOCATION: (378)..(1665)
; OTHER INFORMATION: ORF
; FEATURE:
; NAME/KEY: misc RNA
; LOCATION: (1747)..(2553)
; OTHER INFORMATION: ORF
; FEATURE:
; NAME/KEY: misc RNA
; LOCATION: (2593)..(4011)
; OTHER INFORMATION: ORF
; FEATURE:
; NAME/KEY: misc RNA
; LOCATION: (4013)..(4999)
; OTHER INFORMATION: ORF
; FEATURE:
; NAME/KEY: misc RNA
; LOCATION: (5071)..(6085)
; OTHER INFORMATION: ORF
; US-09-029-603-4

Query Match      4.1%; Score 49.6; DB 3; Length 6085;
Best Local Similarity 44.5%; Pred. No. 2.9e-05;
Matches 196; Conservative 0; Mismatches 244; Indels 0; Gaps 0;

QY 546 CGACAGCTTCCCAATGCTGATGCGGTGATTAAGACGTCGCGTCAATCCGGGATC 605
Db 989 CCAGGAGTTCACCGGCTACTTCGGCGAGAGTGGACCGCGCGGCGGACGACCGGA 1048
QY 606 AGATTTAATTCGATCCTATGTAATCTCTGAATATGAAGCATGCGCGCTCGGACAAAGA 665
Db 1049 CGATCTGCTACCCCTCTGTCGCGCCCGGGACACCGGATCACCGCTCAGCGTGAACGG 1108
QY 666 TATATCGCACTGATTTCAATGCTGTTAGCGCAACGGAACCGGCTGATAAGACGCT 725
Db 1109 CATGTCGGCACTGCGTCACTGCTCACCGCGGCGGACGAGACACCACTGCTCT 1168
QY 726 GCACATGATGATCACTTCTCAATCTCTGACGAGATGATGATGTTTGGCTGA 785
Db 1169 CCGCAGGGGGTCTCACCCTGCGGCCACCGTACGCTCTCGAGAGCTGCGCACAC 1228
QY 786 CGGTTCTGTTAGTTCGAGAGCATTGCGGAGACATTCGCTTATAAACCGCGGTTCA 845
Db 1229 ACGGAGTCGACACCGCGCGGTCGAAGAGCTGATCGGTACGACCCCGCGTCA 1288
QY 846 GATTCGCGGCGAGCTGTCCCAAGATACAGTGTGCGCGGTATGGAATCAAAAAAGATAC 905
Db 1289 GGTGACCGCTGGCGGTAGAGAGACATCCGGCTCGCGGACACGACATCCCGCGG 1348
QY 906 GATTGTTTTTGTATGATCGGTGCGGCTAACCGGACCCCTGAAGCATTTGAACAGCTGA 965
Db 1349 CCGGTTGTCGCGCTGCTCGGCTCGCGGACCGGACCGGCGGCTTCCCGCTCCGA 1408
QY 966 CGTGTTTAATATTATTCATCGG 985
Db 1409 CGTGTGAGCTCCACCGCG 1428

RESULT 7
US-09-724-797-35
; Sequence 35, Application US/09724797
; Patent No. 6733998
; GENERAL INFORMATION:
; APPLICANT: Jon S. THORSON
; TITLE OF INVENTION: MICROMONOSPORA ECHINOSPORA GENES
; TITLE OF INVENTION: ENCODING FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: CALICHEAMICIN AND SELF-RESISTANCE THERETO
; FILE REFERENCE: 2653-40
; CURRENT APPLICATION NUMBER: US/09/724,797
```

```
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 60/111,325
; PRIOR FILING DATE: 1998-12-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 1209
; TYPE: DNA
; ORGANISM: Bacteria
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)....(1209)
; US-09-724-797-35

Query Match      3.9%; Score 47.6; DB 4; Length 1209;
Best Local Similarity 53.8%; Pred. No. 5e-05;
Matches 98; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 805 GCATTGGGAGACATTCGTTTATAACCGCGGTTACGCTGATTCGCGGCGAGCTGTC 864
Db 820 GCCGTGAGGAGTGCCTGCGGTACGACCGCTCGGTGCACTCCAAACACCGCGAGCTCGAC 879
QY 865 CAAGATACAGTGTGCGGCGTATGGAATCAAAAAAGATACGATTCGTTTTTGTATGATC 924
Db 880 GTCGAGCTGAGCTGCGCGGTGCGGCGTGCAGCGACGCTGCTGCTGCTGCGG 939
QY 925 GGTGCGGCTAACCGGAGACCCCTGAAGCATTTGAACAGCCTGAGCTGTTTAAATATTCATCGG 984
Db 940 GCGCGGCGGAACCGGAGACCCCGCGGTCAGACCGCGCGGACGATTCGACATCGAGCGG 999
QY 985 GA 986
Db 1000 GA 1001

RESULT 8
US-08-676-818-1
; Sequence 1, Application US/08676818
; Patent No. 6057136
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley Grant
; APPLICANT: Perkins, John B.
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Pero, Janice G.
; TITLE OF INVENTION: BIOTIN BIOSYNTHESIS IN BACILLUS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,818
; FILING DATE: 08-JUL-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/239,430
; FILING DATE: May 6, 1994
; APPLICATION NUMBER: 08/084,709
; FILING DATE: June 25, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Freeman, John W.
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 04599/004001
; TELECOMMUNICATION INFORMATION:
```

; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8478
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-678-818-1

Query Match 3.7%; Score 44.6; DB 3; Length 8478;
Best Local Similarity 49.4%; Pred. No. 0.0015;
Matches 116; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 744 TTGTCTCAACATCTGACGACATGATGATGTTTGGCTGACCGTTCGTTAGTTCGAG 803
DB 6233 TCTGCTGCAGCATCCAGAACAGCTTTTGAACCTGAGAGAAATCCAGATCTTATTGGTAC 6292

QY 804 AGCCATTGCGGAGACATTCGGTTATAAACCGCGGTTAGCTGATTCGCGGAGCTGTC 863
DB 6293 CCGAGTCGAGGAATGTTTACGCTATGAAAGCCCGCAATATGACAGCCAGAGTTGCGTC 6352

QY 864 CCAAGATACAGTGGTGGCGGTATGGAATCAAAAAGATACGATTTGTTTGTATGAT 923
DB 6353 AGAGGATATTGACATCTGCGGGTGACGATCCGTCAAGGAGAACAGTCTATCTTTTGT 6412

QY 924 CGGTGGGCTAACCGGACCTGAGCATTTGAACAGCCTGACGTTGTTTAATATT 978
DB 6413 AGGAGCGGCTAATCGAGACCTAGCATATTACGAACCCCGATGTTTCGATATT 6467

RESULT 9

US-09-407-549-1
; Sequence 1, Application US/09407549
; Patent No. 6303377
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley Grant
; APPLICANT: Perkins, John B.
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Pero, Janice G.
; TITLE OF INVENTION: BIOTIN BIOSYNTHESIS IN BACILLUS
; TITLE OF INVENTION: SUBTILIS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/407,549
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/239,430
; FILING DATE: May 6, 1994
; APPLICATION NUMBER: 08/084,709
; FILING DATE: June 25, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Freeman, John W.
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 04599/004001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8478
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-09-407-549-1

Query Match 3.7%; Score 44.6; DB 3; Length 8478;
Best Local Similarity 49.4%; Pred. No. 0.0015;
Matches 116; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 744 TTGTCTCAACATCTGACGACATGATGATGTTTGGCTGACCGTTCGTTAGTTCGAG 803
DB 6233 TCTGCTGCAGCATCCAGAACAGCTTTTGAACCTGAGAGAAATCCAGATCTTATTGGTAC 6292

QY 804 AGCCATTGCGGAGACATTCGGTTATAAACCGCGGTTAGCTGATTCGCGGAGCTGTC 863
DB 6293 CCGAGTCGAGGAATGTTTACGCTATGAAAGCCCGCAATATGACAGCCAGAGTTGCGTC 6352

QY 864 CCAAGATACAGTGGTGGCGGTATGGAATCAAAAAGATACGATTTGTTTGTATGAT 923
DB 6353 AGAGGATATTGACATCTGCGGGTGACGATCCGTCAAGGAGAACAGTCTATCTTTTGT 6412

QY 924 CGGTGGGCTAACCGGACCTGAGCATTTGAACAGCCTGACGTTGTTTAATATT 978
DB 6413 AGGAGCGGCTAATCGAGACCTAGCATATTACGAACCCCGATGTTTCGATATT 6467

RESULT 10

US-09-105-537-21
; Sequence 21, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438U1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 1209
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-21

Query Match 3.5%; Score 43; DB 3; Length 1209;
Best Local Similarity 46.5%; Pred. No. 0.0016;
Matches 139; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

QY 746 TGCTCAACATCTCTGACGACATGATGATGTTTGGCTGACCGTTCGTTAGTTCGAGAG 805
DB 770 TCTCGCGCATCCCGAGCAGTGGCGGAGCTGTGCGCCGCGGCTCGCGGCGCG 829

QY 806 CCATTGCGGAGACATTCGGTTATAAACCGCGGTTACGCTGATTCGCGGAGCTGTCCC 865
DB 830 CGGTGGAGGAGACCTCTCGCTACGACCCCGCGGTGCGCTCGACGCCCGGTGTCGG 889

QY 866 AGATACAGTGGTGGCGGTATGGAATCAAAAAGATACGATTTGTTTGTATGATCG 925
DB 890 GGGAGACGAGCTGGCGGCGCGGCTGCCCGCGGGCGCATGTCGTCCTGACCG 949

QY 926 GTGCGGCTAACCGGACCTCTGAAGCATTTGAACAGCCTGACGTTGTTTAATATTTCATCG 985
DB 950 CGCGACCGCGCGGACCCCGGAGGTTCTTACGGAACCGGAGCGCTTCGACCTCGCGCG 1009

QY 986 AAGATCTTGGTATCAAGAGCGCTTTTAGCGGCGCGCGCGCGGATCTCGCTTCGGATCC 1044
DB 1010 CCGACGCGCGCGGACCTCGCGCTGCAACCCCGCGGTCCGTACGCGCCCGGTGCGCTCC 1068

```
RESULT 11
US-09-105-537-3
; Sequence 3, Application US/09105537A
; Patent No. 6285202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 13613
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-3

Query Match      3.5%; Score 43; DB 3; Length 13613;
Best Local Similarity 46.5%; Pred. No. 0.0068;
Matches 139; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

QY 746 TGCCTCAACAATCTGAGCAGATGAATGATGTTTGGCTGACCGTTCGTTAGTTCGAGAG 805
Db 1575 TCCTCGGCATCCGAGCAGTGGCGGAGCTGTGCGACCGCCCGGCTCGCGGCGCC 1634
QY 806 CCATTGCGGAGACATTGCGTTATAAACCGCGGTTTCAGCTGATTCGCGGCGAGCTGTC 865
Db 1635 CGGTGAGGAGACCTCCCTACGACCCCGCGTGCAGTCGACGCGCGGCTGGTCCGCG 1694
QY 866 AGATACAGTGTGCGCGGTATGAAATCAAAAAAGATACGATGTTTGTATGATCG 925
Db 1695 GGGAGACGAGTGGCGGCGCGCGCTGCGCGCGCGCGCGATGTCGTCCTGACCG 1754
QY 926 GTGCGCTAACCGGACCTCGAAGCATTTGAACAGCGCTCAGCTGTTTAATATTCATCGG 985
Db 1755 CCGGACCGCGCGGACCGCGAGGCTTTCAGGACCGCGGAGCGCTCGACCTCGCGCGCC 1814
QY 986 AAGATCTTGATCAAGAGCGCTTTTAGCGGCGCGCGCGCGCATCTCGCTTCGGATCC 1044
Db 1815 CCGACGCGCGCGCACCTCGCGCTGCACCCCGCGGTCCGTACGCGCGCGGTGCGTCC 1873

RESULT 12
US-09-320-878-19
; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
```

```
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-19

Query Match      3.5%; Score 43; DB 3; Length 38506;
Best Local Similarity 46.5%; Pred. No. 0.013;
Matches 139; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

QY 746 TCCTCAACAATCTGAGCAGATGAATGATGTTTGGCTGACCGTTCGTTAGTTCGAGAG 805
Db 35572 TCCTCGGCATCCGAGCAGTGGCGGAGCTGTGCGACCGCGCGGCTCGCGGCGCCG 35631
QY 806 CCATTGCGGAGACATTGCGTTATAAACCGCGGTTTCAGCTGATTCGCGGCGAGCTGCC 865
Db 35632 CGGTGAGGAGACCTTCGCTACGACCCCGCGGTGCGACTCGACGCGCGGTTGTCGCG 35691
QY 866 AAGATACAGTGTGCGCGGTATGAAATCAAAAAAGATACGATGTTTGTATGATCG 925
Db 35692 GGGAGACGAGTGGCGGCGCGCGCTGCGCGCGCGGCGCATGTCGTCCTGACCG 35751
QY 926 GTGCGGCTAACCGGACCTCGAAGCATTTGAACAGCTGACGCTGTTTAATATTCATCGG 985
Db 35752 CCGGACCGCGCGGACCGCGAGGCTTTCAGCGACCGCGGAGCGCTCGACCTCGCGCG 35811
QY 986 AAGATCTTGATCAAGAGCGCTTTTAGCGGCGCGCGCGCATCTCGCTTCGGATCC 1044
Db 35812 CCGACGCGCGCGCACCTCGCGCTGCACCCCGCGGTCCGTACGCGCGCGGTGCGTCC 35870

RESULT 13
US-09-141-908-1
; Sequence 1, Application US/09141908
; Patent No. 6503741
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a
; TITLE OF INVENTION: Modular PKS Gene Cluster as Scaffold
; FILE REFERENCE: 300622002100
; CURRENT APPLICATION NUMBER: US/09/141,908
; CURRENT FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: PROV. 60/076,919
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: PROV. 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-141-908-1

Query Match      3.5%; Score 43; DB 4; Length 38506;
Best Local Similarity 46.5%; Pred. No. 0.013;
Matches 139; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

QY 746 TCCTCAACAATCTGAGCAGATGAATGATGTTTGGCTGACCGTTCGTTAGTTCGAGAG 805
Db 35572 TCCTCGGCATCCGAGCAGTGGCGGAGCTGTGCGACCGCGCGGCTCGCGGCGCCG 35631
```

QY 806 CCATTGCGGAGACATTGCGTTTAAACCGCCGGTTTCAAGTGTGATTCGCGCGAGCTGTCC 865
Db 35632 CGGTGAGGAGACCTTCGCTACGACCCCGCGGTGCGAGTCGACGCGCCGGTGTCCGG 35691
QY 866 AAGATACAGTGTGCGCGGTATGGAATCAAAAAGATACGATGTTTGTATGATCG 925
Db 35692 GGGAGACGAGCTGGCGGCGCGGCTGCCGCGCGCGGTGCTGCTGCTGACCG 35751
QY 926 GTGCGGCTAACCGGACCTGAAGCATTTGAACAGCTCAGCTGTTTAAATATTCATCGG 985
Db 35752 CCGGACCGCGCGGACCCGAGGCTTTCAGGACCCGAGCGCTTCGACCTCGCGCGCC 35811
QY 986 AAGATCTTGTATCAAGAGCGCTTTTTCAGCGCGCGCGCGCATCTCGCTTTCGGATCC 1044
Db 35812 CCGACCGCGCGGACCTCGGCTGCACCCCGCGGTCCGTACGCGCGGTGGCGTCC 35870

RESULT 14

US-09-657-440-19
; Sequence 19, Application US/09657440
; Patent No. 6509455
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30622002120
; CURRENT APPLICATION NUMBER: US/09/657,440
; CURRENT FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: 09/320,878
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: CIP OF 09/141,908
; PRIOR FILING DATE: 1998-08-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-657-440-19

Query Match 3.5%; Score 43; DB 4; Length 38506;
Best Local Similarity 46.5%; Pred. No. 0.013;
Matches 139; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

QY 746 TGCTCAACAATCCTGAGCAGATGAATGATGTTTGGCTGACCGTTGTTAGTTCCGAGAG 805
Db 35572 TCCTCGGCATCCGAGCAGTGGCGGAGCTGTGCGACCGCGCGGCTCGCGGCGCG 35631
QY 806 CCATTGCGGAGACATTGCGTTTAAACCGCGGTTTCAAGTGTGATTCGCGGCGAGCTGTCC 865
Db 35632 CCGTGGAGGAGACCTTCGCTACGACCCCGCGGTGCGATCGACGCGCGGTTGTCGCG 35691
QY 866 AAGATACAGTGTGCGCGGTATGGAATCAAAAAGATACGATGTTTTCATATCATCG 925
Db 35692 GGGAGACGAGCTGGCGGCGCGGCTCCCGCGCGCGGCGCATGTCGTCCTGACCG 35751
QY 926 GTGCGGCTAACCGGACCTTGAAGCATTTGAACAGCTCAGCTGTTTAAATATTCATCGG 985
Db 35752 CCGGACCGCGCGGACCCGAGGCTTTCACGACCCGAGCGCTTCGACCTCGCGCGCC 35811
QY 986 AAGATCTTGTATCAAGAGCGCTTTTTCAGCGCGCGCGCGCATCTCGCTTTCGATCC 1044
Db 35812 CCGACGCGCGCGGACCTTCGCTGCACCCCGCGGTCCGTACGCGCGGTGGCGTCC 35870

RESULT 15

US-08-576-626A-1
; Sequence 1, Application US/08576626A
; Patent No. 5998194
; GENERAL INFORMATION:

; APPLICANT: Summers, R.G.
; APPLICANT: Katz, L.
; APPLICANT: Donadio, S.
; APPLICANT: Staver, M.J.
; TITLE OF INVENTION: POLYKETIDE-ASSOCIATED SUGAR
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/576,626A
; FILING DATE: 21-DEC-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dianne Casuto
; REGISTRATION NUMBER: P-40,943
; REFERENCE/DOCKET NUMBER: 5857.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (847) 938-3137
; TELEFAX: (847) 938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3756 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-576-626A-1

Query Match 3.5%; Score 42.4; DB 2; Length 3756;
Best Local Similarity 47.1%; Pred. No. 0.0049;
Matches 130; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 691 CTGTTAGCCGCAACCGAACCGGCTGATGAAGACGCTGGCACTGATGATCTACCATTTGCTC 750
Db 648 CTGTTGCGCGGCGGCGAGATGACCGCGAACCGTGGTGCAGCGGTCCTGGCGGTCG 707
QY 751 AACATCTCTGACGAGATGAATGATGTTTGGCTGACCGTTTGTAGTTCGAGAGCCATT 810
Db 708 GCCGAACCGGGGCTGGCGGAACGGATCGCCGACGACCCCGCGCGCGAGAACCGTTC 767
QY 811 GCGGAGACATTGCGGTTATAAACCGCGGTTTACGCTGATTTCGCGCGAGCTGCCCAAGAT 870
Db 768 GCGGAGGTCGTGCGGCTGCACCGCGCATTCACCTGGAGCGCGCACCGCCACCGCAGAG 827
QY 871 ACAGTGTGCGCGGTATGGAATCAAAAAGATACGATGTTTGTATGATCGGTGCG 930
Db 828 GTGCGGCTCGCGAGCAGCTGATCGCGAAGCGGAGGAGTCTGTGTCGTGCGCGCG 887
QY 931 GCTAACCGGGACCTTGAAGCATTTGAACAGCTGAC 966
Db 888 GCCAACCGCGACCCGAGGTCCTTCGCGAGCCCGAC 923

Search completed: December 10, 2004, 18:17:12
JOB time : 159.858 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2004, 11:49:51 ; Search time 5403.29 Seconds
(without alignments)
10633.698 Million cell updates/sec

Title: US-10-627-124-1
Perfect score: 1215
Sequence: 1 atgagccaatgattaaatt.....tcgttgcttgacggggca 1215

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_hg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1215	100.0	47739	1	AF017113	AF017113 Bacillus
c 2	1215	100.0	194692	1	BSUB0018	Z99121 Bacillus su
3	617.2	50.8	1208	6	AX433766	AX433766 Sequence
c 4	247.4	20.4	349483	1	BX571859	BX571859 Photobab
5	247.4	20.4	349980	6	AX770907	AX770907 Sequence
6	171.8	14.1	36888	1	AY260760	AY260760 Streptomy
7	169	13.9	314100	1	SC0939106	AL939106 Streptomy
c 8	116.2	9.6	300885	1	AE012559	AE012559 Xylella f
9	113.6	9.3	10092	1	AE003889	AE003889 Xylella f
c 10	90.8	7.5	30090	1	AE017008	AE017008 Bacillus
11	89.4	7.4	10261	1	AE003887	AE003887 Xylella f
12	88.8	7.3	190050	1	AL646080	AL646080 Ralstonia
c 13	86.8	7.1	300885	1	AE012559	AE012559 Xylella f
c 14	85.2	7.0	302325	1	AE017236	AE017236 Mycobacte
15	78.8	6.5	35739	1	AF015825	AF015825 Bacillus
16	78.8	6.5	201241	1	BSUB0007	Z99110 Bacillus su
17	75.2	6.2	110000	1	CR543861	Continuation (16 o
18	72.4	6.0	966	6	AX433775	AX433775 Sequence
c 19	69.4	5.7	39228	1	MLCB1788	AL008609 Mycobacte

c 20	69.4	5.7	342300	1	MLEPRTN8	AL583924 Mycobacte
21	68.2	5.6	9180	1	AF274704	AF274704 Pseudomon
c 22	67.8	5.6	110000	1	AE000516_39	Continuation (40 o
c 23	67.8	5.6	349606	1	BX842583	BX842583 Mycobacte
24	67.6	5.6	299925	1	AP005039	AP005039 Streptomy
c 25	67.2	5.5	3187	1	BACRTP	M24523 B.subtilis
26	67.2	5.5	300327	1	AE017228	AE017228 Mycobacte
27	67.2	5.5	300893	1	AE017006	AE017006 Bacillus
28	64.8	5.3	1213	6	AX433767	AX433767 Sequence
c 29	64	5.3	300327	1	AE017228	AE017228 Mycobacte
c 30	63.2	5.2	316050	1	BX248346	BX248346 Mycobacte
c 31	61.8	5.1	110000	1	AE000516_08	Continuation (9 of
c 32	61.8	5.1	320050	1	BX248336	BX248336 Mycobacte
c 33	61.8	5.1	349564	1	BX842574	BX842574 Mycobacte
34	61.2	5.0	32495	1	AF329849	AF329849 Agrobacte
c 35	61.2	5.0	75778	1	AY328003821	AY328003 Symbiont
36	59.6	4.9	309267	1	AE017235	AE017235 Mycobacte
c 37	59.4	4.9	26386	1	AF119621	AF119621 Pseudomon
38	59.2	4.9	17512	1	AB071405	AB071405 Lechevali
39	59.2	4.9	25681	1	SAB414559	AJ414559 Saccharot
40	59.2	4.9	26144	1	AB090952	AB090952 Lechevali
c 41	59.2	4.9	28654	1	AF534707	AF534707 Lechevali
42	58.4	4.8	8407	1	SEU82823	U82823 Saccharopol
43	58.4	4.8	313846	1	AE017234	AE017234 Mycobacte
44	58.2	4.8	128136	1	AF440524	AF440524 Pseudomon
45	58.2	4.8	302325	1	AE017236	AE017236 Mycobacte

ALIGNMENTS

RESULT 1	AF017113	Bacillus subtilis	47739 bp	DNA	linear	BCT 23-MAY-2000
LOCUS	AF017113	Bacillus subtilis	300-304	degree	genomic	sequence.
DEFINITION	AF017113	Bacillus subtilis	300-304	degree	genomic	sequence.
ACCESSION	AF017113.1	GI:2618830				
VERSION	AF017113.1	GI:2618830				
KEYWORDS		Bacillus subtilis				
ORGANISM		Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.				
REFERENCE	1 (bases 1 to 47739)	Reizer, J., Hoischen, C., Titgemeyer, F., Rivolta, C., Rabus, R., Stulke, J., Karamata, D., Sailer, M. H. Jr. and Hillen, W.				
AUTHORS		A novel protein kinase that controls carbon catabolite repression in bacteria				
TITLE		Mol. Microbiol. 27 (6), 1157-1169 (1998)				
JOURNAL		98230327				
MEDLINE		9570401				
PUBMED		2 (bases 30303 to 37035)				
REFERENCE		Robinson, C., Rivolta, C., Karamata, D. and Moir, A.				
AUTHORS		The product of the yvoC (grf) gene of Bacillus subtilis is required for spore germination				
TITLE		Microbiology 144 (Pt 11), 3105-3109 (1998)				
JOURNAL		99061203				
MEDLINE		3 (bases 1 to 47739)				
PUBMED		Lazarevic, V., Soldo, B., Rivolta, C., Reynolds, S., Maue, C. and Karamata, D.				
REFERENCE		Nucleotide sequence of the 300-304 chromosomal segment of Bacillus subtilis				
AUTHORS		Unpublished				
TITLE		4 (bases 1 to 47739)				
JOURNAL		Lazarevic, V.				
REFERENCE		Direct Submission				
AUTHORS		Submitted (05-AUG-1997) IGBM, Cesar-Roux 19, Lausanne 1005, Switzerland				
TITLE		Location/Qualifiers				
JOURNAL		1. .47739				
REFERENCE		/organism="Bacillus subtilis"				
AUTHORS		/mol_type="genomic DNA"				
TITLE		/db_xref="taxon:1423"				
JOURNAL		/map="300-304 degrees"				

gene	71. .1184 /gene="prfB"	gene	ADPTGNLDPDTSWEVMKLTBEEINNRTTVMVATHNKIEIVNTMKRKRVAIEDGIIVRD ESRGEYGSYD"
RBS	71. .77 /gene="prfB"	RBS	3476. .4385 /gene="ftsX"
CDS	Join(83. .154,156. .1184) /gene="prfB" /note="at low concentrations of RF-2 nucleotides 155. .158 (UGAC) are translated as Asp, Phe, H.J. Nucleic Acids Res. 20. 4423-4428, 1992" /codon_start=1 /transl_table=11 /product="putative peptide chain release factor RF-2" /protein_id="AAC67303.1" /db_xref="GI:2618874" /translation="WELSIKRAELNWAASRLADFRGSLDLESKEARIAELDEQMAADPE FNDQQAQTVINEANGSLDYNNYKKNLNEHEELQMTDHLKBPDTDLQLELSKEL KSLTEFNEFEFLQLLSPEYDKNNAILLELPGAGGTESQDWMGSMLLRMYTRWGERRGF KYETLDYLPDGDAGIKSVTLIKGHNAVGYLKAEGVHRLVRIISPDSSGRRTTSFVS CBWMPFENDEIDIRTEDIKVDITYRASGAGGQHVNTDSAVRIIHLPTNVVVTQOTE RSQIKNRERAMKWLAKLYORBIERQQAELEDRGEQKEIGWGSQIRSVVFPHPYSVK DHRVTENGVQAVMDGDDITFDIAYLSKLS"	CDS	3476. .3484 /gene="ftsX"
	143. .149 /gene="prfB"		3495. .4385 /gene="ftsX"
RBS	1338. .2195 /gene="yvJ"	RBS	4511. .5967 /codon_start=1 /transl_table=11 /product="cell division protein"
gene	1338. .1342 /gene="yvJ"	gene	4511. .5967 /db_xref="GI:2618835"
RBS	1350. .2195 /gene="yvJ"	RBS	4525. .5967 /gene="yvJ"
CDS	1350. .2195 /note="similar to thermophilic bacterium PS-3 hypothetical protein: PIR Accession Number S43727 and to Bacillus subtilis YqfU, YpJc and YitB proteins"	CDS	4525. .5967 /codon_start=1 /transl_table=11 /product="putative protease"
	143. .149 /product="YvJ"		4511. .5967 /protein_id="AAC67263.1"
	143. .149 /db_xref="GI:2618831"		4511. .5967 /db_xref="GI:2618834"
	143. .149 /translation="WDVRNKTLLIIRDVYVILIGAITAVSFNVFLPNKIAAGGVSG ISTILQSGFEAYVQVNIINIEPLIAGVILGKGFGKTLGSLVFLPLVFLVTRDIQP ATHHELLAIFGSGVIGIGIVILGKSTGTGTALAAQIIHKYSLGSLGKCLAIIDGM IYVTAVIENIQGLYAMLVGVYSSKTDVVGQFNRSKMLIITKQSQAVKEAVLQK IDRGVTKISAVGGYTDORPILMVCVVGQTEFTKLQIVKQIDESAFAVIVADASEVLGE GPKRA"		4511. .5967 /translation="MNOKIMAVIAAGSMFGAGVYAGINLEMDKDPOTAAVPATAQA DSEDRKAMDKTEKAYELISNEYVEKVDREKLEGAIOGLMLSTLNDPPYSVYMDKOTAKO PSDLSSFEIGIABGVCHEDGKIIIVSPFKSPAKAGLKNDEILISINGSMAKD L NHAULIRKRGSSVSMKIQRPCTKQSLFRIKRAEIPLETVFASEKKVQGSVGYIA ISTFSEHTAEDFAKALRELEKEIEGLVIDVRGNPGGYLQSVVEILHFVTKDQPYIQ IAEKRGDKRYFSTLTHKKYAPVNVITDKGSASASEILAGALKEAGHYDVVGDTSFGK GTVQOAVPMGDGSKI LTYKWLTPNGNWIHKGIEPTIATIKQPYFSAGPLQKPEPL KVDNNEDVHAQVLLKGLSPDRGEDGYESDKMKKAVMAFQDQNKLNKTGVIDTRTA ETLNQOIEKKSKDENLQLOTALKSLFVN"
gene	2223. .2582 /gene="cccB"	gene	6773. .7977 /gene="yvJ"
RBS	2223. .2229 /gene="cccB"	RBS	6773. .6777 /gene="yvJ"
CDS	2244. .2582 /gene="cccB"	CDS	6784. .7977 /gene="yvJ"
	2223. .2582 /note="putative lipoprotein"		6784. .7977 /note="transmembrane protein"
	2223. .2582 /evidence="experimental"		6784. .7977 /codon_start=1 /transl_table=11 /product="YvJ"
	2223. .2582 /transl_table=11 /product="cytochrome c-551"		6784. .7977 /protein_id="AAC67265.1"
	2223. .2582 /protein_id="AAC67261.1"		6784. .7977 /db_xref="GI:2618836"
	2223. .2582 /db_xref="GI:2618832"		6784. .7977 /translation="MSVQVQIELLKLSAGLFFHLFWFFIITLAFGVVRIKRRKTF HTRIADYDDLTFTYTKGLIPLGLLIVGLGIGISIPGLLAIIVATAAATLIRAN WMSAAYIVSVMLIGFGLQIYQAEFFLERFPGQFVAVMPAVAFVGLGLIITTEGAVYR SAHVRTSPALVSVSRGLPIGQOLANRWLLPLFLVPGNGLSHLSWMPVFTVPGGSF HELMIPVFGQVRQVQSLPETSIRITAKRVCLGLAVLGAVALSLLMTPLAGAUCT ALLGRIFLSIKQRVNDNAPYFPRKRDQGLMVGLIIPNTPAEDLEKIGEITKNGI PVKVSQDPYEQALQHNRAVYKLEIIGLNGEIRFDQKASYEGERHEHGLILFVKDDRDEA VASQV"
	2223. .2582 /translation="MKSLSILMIGFALSIVLLAAGSNDAAKEKTDTSKTEATASEG BELYQOSCVGCHGKDLGVSGNPLQVGGKYDEHKIESIINKRGNMFKGLVDDNEAA VIAKWLSEKK"		6784. .7977 complement (8025. .9373) /gene="yvK"
gene	2806. .3502 /gene="ftsE"	gene	complement (8025. .9359) /gene="yvK"
RBS	2806. .2813 /gene="ftsE"	RBS	complement (8025. .9359) /note="belongs to the major facilitator family"
CDS	2816. .3502 /gene="ftsE"	CDS	complement (8025. .9359) /codon_start=1 /transl_table=11 /product="YvK"
	2816. .3502 /codon_start=1 /transl_table=11 /product="cell division ATP-binding protein"		complement (8025. .9359) /product="YvK"
	2816. .3502 /protein_id="AAC67262.1"		complement (8025. .9359) /db_xref="GI:2618837"
	2816. .3502 /db_xref="GI:2618833"		complement (8025. .9359) /translation="MSSRKWALIVSLLGAILVPIINMTIAVALSSISHTYNESIAS ITWVTVLIVMAVTPQIAGKLDGMYGKNTWLMGVGLFIASLGALSPSLLLIVF RALQVGGALLTPNSIAIRHVSEKRLPKVFGFGLGAGLGAALGPTFGISLLIDFS WHSIFWNIPFLAIFALTALTMFPQYKKNKSDAPLDIIGSLLLAGSIVSIILLTKNEA

PMGYTVSVLILFLVPLFRPRBRKTRQHPIDFALPKSFTTNNANISVLLSNLMVAVL	
LIMPLMTQFGLNNSGMSVFSIFMSASNVWGAQJLHKWGAQKIIIFLFAWMAG	
ANLLFLILSSSVLFLMVLISLILGLASVGLTSMQVSLATVDPMCSVGFAGIFSTF	
RVFGSTISSALIGLSVHTLPMILFAVSIIGVFSLIGSIKIDETARIEKNSA	
RBS	complement (9366..9373)
gene	/gene="yvkA"
	complement (9390..9972)
CDS	/gene="yvkB"
	/note="9959"
CDS	complement (9390..9959)
	/gene="yvkB"
CDS	/note="belongs to the TetR/AcrR family of transcriptional regulators"
	/codon_start=1
Query Match 100.0%; Score 1215; DB 1; Length 47739;	
Best Local Similarity 100.0%; Pred. No. 0;	
Matches 1215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 ATGAGCCAATCGATTAAATTTGTTAGTGTCTTTCTGATCAATTTCAAAACAATCCATAT 60
DB	24518 ATGAGCCAATCGATTAAATTTGTTAGTGTCTTTCTGATCAATTTCAAAACAATCCATAT 24577
QY	61 GCTTATTTTTCACACTCGGGAGGAGATCCGGTTCATTATGAGAGTCGATAGACGT 120
DB	24578 GCTTATTTTTCACACTCGGGAGGAGATCCGGTTCATTATGAGAGTCGATAGACGT 24637
QY	121 TATTTTATCAGCGCTATCATGATGTCGGCTATATCTTTTTCAGCATCCGATATCTTTCAG 180
DB	24638 TATTTTATCAGCGCTATCATGATGTCGGCTATATCTTTTTCAGCATCCGATATCTTTCAG 24697
QY	181 ACGAAATCACTTGTGAGCGTCCGACCAAGATCATGCGAGGCGCTGTGCTGCCCAATG 240
DB	24698 ACGAAATCACTTGTGAGCGTCCGACCAAGATCATGCGAGGCGCTGTGCTGCCCAATG 24757
QY	241 CATGGAAGAAGACACTCTGCCAAAGAGATTTGATGAGAGCTTTATCGGTGACGCA 300
DB	24758 CATGGAAGAAGACACTCTGCCAAAGAGATTTGATGAGAGCTTTATCGGTGACGCA 24817
QY	301 CTGGATCATCTGTCTCCATTTGATTAACCAAAATGACAGAAATCTGTAGCGCTTATCTT 360
DB	24818 CTGGATCATCTGTCTCCATTTGATTAACCAAAATGACAGAAATCTGTAGCGCTTATCTT 24877
QY	361 GAAAGAGGAAAGATGATCTCGTCAATGATTTTGGAAAGACGTTTTCGGTGTGCGTCACG 420
DB	24878 GAAAGAGGAAAGATGATCTCGTCAATGATTTTGGAAAGACGTTTTCGGTGTGCGTCACG 24937
QY	421 ATGACATGCTCGGGCTGATATAAGAGACCAATGAAAAAATCTCTGAGTGGACACGCGGA 480
DB	24938 ATGACATGCTCGGGCTGATATAAGAGACCAATGAAAAAATCTCTGAGTGGACACGCGGA 24997
QY	481 GTTGGCGATTTTATCAGAGTATCTCTCAATCTCTGAGCGCGGACCATTCGTTATCG 540
DB	24998 GTTGGCGATTTTATCAGAGTATCTCTCAATCTCTGAGCGCGGACCATTCGTTATCG 25057
QY	541 TGCAGCGAAGACGCTTTCCCAATCTTGTATGTCGGGTCAATTAAGAAAGCTCGCGCAATCG 600
DB	25058 TGCAGCGAAGACGCTTTCCCAATCTTGTATGTCGGGTCAATTAAGAAAGCTCGCGCAATCG 25117
QY	601 GGATCAGATTTAATTTTCGATCTTATGATCTTCTGAAATATGAAGGCGATCGCGTTCGGAC 660
DB	25118 GGATCAGATTTAATTTTCGATCTTATGATCTTCTGAAATATGAAGGCGATCGCGTTCGGAC 25177
QY	661 AGGATATCTCGCACTGATCTTCTATGCTGTGTTAGCGGACCAAGCGCTGTATAG 720
DB	25178 AGGATATCTCGCACTGATCTTCTATGCTGTGTTAGCGGACCAAGCGCTGTATAG 25237
QY	721 ACGCTGGCACTGATGATCTACCAATTTGCTCAAAATCTCTGAGCAGATGAATGATTTTG 780
DB	25238 ACGCTGGCACTGATGATCTACCAATTTGCTCAAAATCTCTGAGCAGATGAATGATTTTG 25297
QY	781 GCTGACCGTTCGTTAGTTCGAGAGCCATTCGGGAGACATTCGTTTATAAACCGCGGTT 840

DB	25298	GCTGACCGTTCGTTAGTTCGAGAGCCATTCGAGAGACATTCGTTATATAACCGCGGTT	25357
QY	841	CAGCTGATTCGCGGACGCTGTCCCAAGATACAGTGGTCCGCGGTATGAAATCAAAAA	900
DB	25358	CAGCTGATTCGCGGACGCTGTCCCAAGATACAGTGGTCCGCGGTATGAAATCAAAAA	25417
QY	901	GATACGATTTGTTTGTATGATTCGTTGCGGTAAACCGGGAACCTGAAGCATTTGAACAG	960
DB	25418	GATACGATTTGTTTGTATGATTCGTTGCGGTAAACCGGGAACCTGAAGCATTTGAACAG	25477
QY	961	CCTGACGTTGTTTAAATATTCATCGGAAGATCTTGTGATCAAGAGCGCTTTAGCGCGCC	1020
DB	25478	CCTGACGTTGTTTAAATATTCATCGGAAGATCTTGTGATCAAGAGCGCTTTAGCGCGCC	25537
QY	1021	GCCCGGCATCTCGCTTTCCGATCCGCGCATTCATAACTGTGTAGGAGCAGCTTTTGCAAA	1080
DB	25538	GCCCGGCATCTCGCTTTCCGATCCGCGCATTCATAACTGTGTAGGAGCAGCTTTTGCAAA	25597
QY	1081	AACGAAATCGAAATTTGTAGCTAATATTTGCTGTGATAGATCGGAATATCAGATTAGAG	1140
DB	25598	AACGAAATCGAAATTTGTAGCTAATATTTGCTGTGATAGATCGGAATATCAGATTAGAG	25657
QY	1141	GAAGATTTTGTATGCTGAGTCCGCTCTGTATACACGCGGACCTTTTCACTTCCTGTT	1200
DB	25658	GAAGATTTTGTATGCTGAGTCCGCTCTGTATACACGCGGACCTTTTCACTTCCTGTT	25717
QY	1201	GGCTTTGACGGGCA 1215	
DB	25718	GGCTTTGACGGGCA 25732	
RESULT 2			
BSUB0018/c			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
AUTHORS			
REFERENCE			
1 (bases 1 to 194692)			
Bacillus subtilis subsp. subtilis str. 168			
Bacillus subtilis subsp. subtilis str. 168			
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
Kunat, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,			
Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S.,			
Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C.,			
Bron, S., Brouillet, S., Bruschi, C.V., Caldwell, B., Capuano, V.,			
Carter, N.M., Choi, S.K., Codani, J.J., Conner, I.F., Cummings, N.J.,			
Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D.,			
Emmerson, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrari, E.,			
Fouger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A.,			
Galleron, N., Ghim, S.Y., Glaser, P., Goffeau, A., Golighly, E.J.,			
Grandi, G., Guiseppe, G., Guy, B.J., Haga, K., Haech, J., Harwood, C.R.,			
Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F.,			
Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y.,			
Klaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P.,			
Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A.,			
Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H.,			
Masuda, S., Maue, C., Medigue, C., Medina, N., Mellado, R.P.,			
Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M.,			
Ogawa, K., Ogiwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M.,			
Portetle, D., Porwollik, S., Prescott, A.M., Presecan, E., Pujic, P.,			
Purtelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M.,			
Rivolta, C., Roche, E., Roche, B., Rose, M., Sadaie, Y., Sato, T.,			
Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J.,			
Sekowska, A., Seror, S.J., Serror, P., Shin, B.S., Soldo, B.,			
Sorokin, A., Taccioni, E., Takagi, T., Takahashi, H., Takemaru, K.,			
Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A.,			
Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A.,			
Viari, A., Wambutt, R., Wedler, H., Wedler, H., Weitzenegger, T.,			
Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K.,			
Yoshida, K., Yoshikawa, H.F., Zumstein, E., Yoshikawa, H. and			
Danchin, A.			

TITLE JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE JOURNAL	The complete genome sequence of the gram-positive bacterium Bacillus subtilis Nature 390 (6657), 249-256 (1997) 98044033 9384377 2 (bases 1 to 194692) Kunét,F., Ogaawara,N., Yoshikawa,H. and Danchin,A. Direct Submission Submitted (27-JUN-2003) I. Moszer, A. Danchin, Institut Pasteur, Genetique des Genomes Bacteriens, 28 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr, adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45 68 89 48 On Jul 7, 2003 this sequence version replaced gi:2635827. This entry contains data from release R16.1 of the Subtilist database. Further data on gene annotation and detailed information about changes from previous releases can be found at http://genolist.pasteur.fr/Subtilist/ .	Location/Qualifiers 1. .194692 /organism="Bacillus subtilis subsp. subtilis str. 168" /mol_type="genomic DNA" /strain="168" /db_xref="taxon:224308" complement(79. .888) /gene="thuC" /locus_tag="BSU33290" complement(79. .888) /gene="thuC" /locus_tag="BSU33290" /codon_start=1 /transl_table=11 /product="ferrichrome ABC transporter (ATP-binding protein)" /protein_id="CAB15335.1" /db_xref="GI:2635843" /db_xref="GOA:P49938" /db_xref="Swiss-Prot:P49938" /translation="WNSLSTELQIGYGDRIVEDLNISIPKGIITLIGNGCGKST ILKTSRIMRSAGAVYNGRAIKHMSKTDIAKDWAILPQPEAPSGLTHVLSYGR FPHQSGFRLNDEDRRIIKWLEETGMAEYARPEALUSGGQRQVRWIAMALAQGTG LLDEPTTDLAQLEILQLDLRLKSGQRTILMVIHDLNHAARFSHYMLKKGTV IKEGTALVMTPTDLKQVFDIAEIVTDPRTNKVCLATYDLIKNEKELVTV" complement(904. .1914) /gene="thuG" /locus_tag="BSU33300" complement(904. .1914) /gene="thuG" /locus_tag="BSU33300" /codon_start=1 /transl_table=11 /product="ferrichrome ABC transporter (permease)" /protein_id="CAB15336.1" /db_xref="GI:2635844" /db_xref="GOA:P49937" /db_xref="Swiss-Prot:P49937" /translation="WKTNTKPLIYVATFLLIVVFTLSNLGVIKIAPLKTQVFF GGTARDSELFEPRLLIISLLVAGISVAGAILQSVQNELAEPGLGNAGGSL AVLTYFFQSGASDLSPFGTFMPLFSALAGAILAFLIYLANKKGVTPRLILVGI WAGNFANLLIIFQKMDPHDFMQAAVWISGSIGWANGNMIWAILPWIVILLFLTYK ARYLNMQLDGLATGLGTAVKERRILLAAVTLAASCVAAGGIAFLGIAPHVAR RLTGPRHQTLIPVSFIGSFLFLADTLARNVLAPSEIPVGLIVISVLGAPYFIYLLMK AN" complement(1914. .3068) /gene="thuB" /locus_tag="BSU33310" complement(1914. .3068) /gene="thuB" /locus_tag="BSU33310" /codon_start=1 /transl_table=11 /product="ferrichrome ABC transporter (permease)" /protein_id="CAB15337.1"	gene CDS gene CDS gene CDS gene CDS gene CDS gene CDS gene CDS	/db_xref="GI:2635845" /db_xref="GOA:P49936" /translation="MHFPCSKHSIKSAEKSILKQQLIIISNKEVRLSOHKNIR TASBEIQWTSRTYGAVIIVLAGLCLGLGAFSLISLGAADIHLRTVWEAITHYQPTKT SHQIHDLRLPRTAAALVLAGLAVSGAIMQGMTRNPLAESIMGVTSQSAFIAF AFFPLSAGMLVMSFAGAGLGAFTVMGIGMFSRGLTPVKLALAGTAVTFFTGSL AIAIRFDVAQDISFWAGVAGVKWSGVQLLLIAGAVGLTAFITARVTVLSLGGDL KAGLGOYTSVAKLVGLMIVLITLGAAVSIAGTIAFIGLIIPHITFLVGVYRWIIPC SAVLGAVLLVPADIAARLVNAPFTFPGALTSLIGVPPFFYLARRRGL" 3166. .4113 /gene="fhud" /locus_tag="BSU33320" 3166. .4113 /gene="fhud" /locus_tag="BSU33320" /codon_start=1 /transl_table=11 /product="ferrichrome ABC transporter (ferrichrome-binding protein)" /protein_id="CAB15338.1" /db_xref="GI:2635846" /db_xref="GOA:P37580" /db_xref="Swiss-Prot:P37580" /translation="WTHYKLGAAFFALLIIAALACNNSEKSGSADSKGAEFTT YKAENGIVKI PKPKRVVMADGYGVFKTLGINVVGAPENVKPNPYKGTNGVNI GDGTSEKVIDLNPDLIIWMTQAGDIKKLEKIAPTVAVKYDKLDNIQLKEPAKWTG TEDRAEKRLAKWDKXAAAKYIKAVGDKTISIMQNGKDIYVFGKDFGRGGSIIYK DLGQATKLTKKKAIDPGYTSISLEKLPDPAGDIYPAGPWQSGDDGGVFESSIWK NLNAVXGHVYKMDPIGIFTDPILEGQLEFITESLTK" complement(4348. .5757) /gene="yvsh" /locus_tag="BSU33330" complement(4348. .5757) /gene="yvsh" /locus_tag="BSU33330" /function="unknown" /note="similar to ABC transporter (amino acid permease)" /codon_start=1 /transl_table=11 /protein_id="CAB15339.1" /db_xref="GI:2635847" /db_xref="GOA:O32204" /db_xref="TrEMBL:O32204" /translation="MEQTKWGFLLTAFFVGNVVGSGIFSLPSSLASIASPFGATSA WLLTGAVLMIALVFGHLSIRKPELTAGPQSYARALSDPKKGNRAGTWMVGYVAS WISNVAITSLAGYLTSPFPLVDKREMFSGGQEVTLGQLTTFVAVCTILLWYTHAIL VASINGASKLNFVTLTKVLGFFVFI VAGLEVFTSLGFHFYFQVGGENGTSIGGGQ VHNAIISTLWAFVGSIAVILSGRARSQDVKRATITGLILALSIIYIIVTITMGVLP HDKLVGSEKPPVDVLYAVGNAGSVIALLAILCLFGTMLGWLIGSVPYQAAGADP PPAFPAKTKKGSPIVAILITNVMSQVPIFSVISRTISDFTLTATLAVLIPLV SAIYSLKVVINGETYDQKGSVRDGLIALLACAYSVPVITGTADTLTFLIGLGF VGLIYPPVSNKFKQEQQA" complement(6157. .6297) /gene="sspu" /locus_tag="BSU33340" complement(6157. .6297) /gene="sspu" /locus_tag="BSU33340" /codon_start=1 /transl_table=11 /product="small acid-soluble spore protein" /protein_id="CAE01467.1" /db_xref="GI:32468826" /translation="MGPFNDKDKRSEKKNVIQGALEBAGSALKDDPLQEAIVQKKXN NR" 6464. .6946 /gene="yvsG" /locus_tag="BSU33350" 6464. .6946 /gene="yvsG" /locus_tag="BSU33350" /function="unknown"
---	--	---	---	---

/note="similar to unknown proteins"	
/codon_start=1	
/transl_table=11	
/protein_id="CAB15340.1"	
/db_xref="GI:2635848"	
/db_xref="GOA:032205"	
/db_xref="Swiss-Prot:032205"	
/translation="MTGTHIMGSIASCTAAAYYYGDFPVLMAASGAVGALIPDICTH QSKIGKFPILSKVSVFGRHTFTSLFLMIFPITSTYIPDNISAGLMIGMASH LILDATVNGIKLLPSTIRVRLPLYMKTGPSFQSLVLAGLTLASCYVYMLFHGRMF "	
gene	7046..8899
/gene="yvgJ"	
/locus_tag="BSU33360"	
Query Match	100.0%; Score 1215; DB 6; Length 194692;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 1215; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 ATGAGCCAATCGAATTAATGTTTGTAGTGTCTTCTGATCAATTTCAAAACAATCCATAT 60
DB	188497 ATGAGCCAATCGAATTAATGTTTGTAGTGTCTTCTGATCAATTTCAAAACAATCCATAT 188438
QY	61 GCTTATTTTCAACTGCGGAGGAAGATCGGTTTCAATATGAAGATCGATAGACAT 120
DB	188437 GCTTATTTTCAACTGCGGAGGAAGATCGGTTTCAATATGAAGATCGATAGACAT 188378
QY	121 TATTTTATCAGCGCTATCATATGTCGCTATATCTTTCAGATCCGATATCTTCACG 180
DB	188377 TATTTTATCAGCGCTATCATATGTCGCTATATCTTTCAGATCCGATATCTTCACG 188318
QY	181 ACGAATACCTTGTGAGCGTCCGACCAAGATCAGCGAGCCCTGTGTCGCCCAATATG 240
DB	188317 ACGAATACCTTGTGAGCGTCCGACCAAGATCAGCGAGCCCTGTGTCGCCCAATATG 188258
QY	241 CATGGAAGAAGACACTCTGCCAAAAGAAATTTAGTGAAGACCTTTATCGGTGACGCA 300
DB	188257 CATGGAAGAAGACACTCTGCCAAAAGAAATTTAGTGAAGACCTTTATCGGTGACGCA 188198
QY	301 CTGGATCATCTGCTCCATTTGATTAACAAAATGACAGAAACTTTGTAGCGCTTATCTT 360
DB	188197 CTGGATCATCTGCTCCATTTGATTAACAAAATGACAGAAACTTTGTAGCGCTTATCTT 188138
QY	361 GAAAGAGGAAAGATGATCTCGTCAATGATTTTGGAAAGACGTTTTCGGTGTGCGTCACG 420
DB	188137 GAAAGAGGAAAGATGATCTCGTCAATGATTTTGGAAAGACGTTTTCGGTGTGCGTCACG 188078
QY	421 ATGGACATGCTCGGGCTGGATTAAGAGACCATGAAATAATCTCTGAGTGGCACAGCGGA 480
DB	188077 ATGGACATGCTCGGGCTGGATTAAGAGACCATGAAATAATCTCTGAGTGGCACAGCGGA 188018
QY	481 GTTCCGATTTTATCAGAGTATCTCTCAATCTCTGAGCGCGGCACATTCGTTATCG 540
DB	188017 GTTCCGATTTTATCAGAGTATCTCTCAATCTCTGAGCGCGGCACATTCGTTATCG 187958
QY	541 TGCAGCGAAGACGCTTTCCCAATCTTGTATGTCGGGTCAITTAAGAAACGTCGCGTCAATCG 600
DB	187957 TGCAGCGAAGACGCTTTCCCAATCTTGTATGTCGGGTCAITTAAGAAACGTCGCGTCAATCG 187898
QY	601 GGATCAGATTTAATTTTCGATTCCTATGATCTTCTGAATATGAAGGATGCGCTGTCGAC 660
DB	187897 GGATCAGATTTAATTTTCGATTCCTATGATCTTCTGAATATGAAGGATGCGCTGTCGAC 187838
QY	661 AGGATATCTCGACTGCTTCTTATGTCGTGTAGCCGCAACGCAACGCGCTGATAAG 720
DB	187837 AGGATATCTCGACTGCTTCTTATGTCGTGTAGCCGCAACGCAACGCGCTGATAAG 187778
QY	721 ACCTGGCACTGATGATCTACCAATTTGCTCAACAATCTCTGAGCAGATGAATGATGTTTG 780
DB	187777 ACCTGGCACTGATGATCTACCAATTTGCTCAACAATCTCTGAGCAGATGAATGATGTTTG 187718
QY	781 GCTGACCGTTCGTTAGTTCGAGAGCAATTCGTTATTAACCGCGCGTT 840

RESULT 3

AX433766

LOCUS

Sequence 2181 from Patent WO0229113.

AX433766

AX433766

AX433766.1

GI:21658574

KEYWORDS

ORGANISM

Bacillus licheniformis

Bacillus licheniformis

Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

REFERENCE

1

Berka, R. and Clausen, I. G.

Methods for monitoring multiple gene expression

Patent: WO 0229113-A 2181 11-APR-2002;

Novozymes Biotech, Inc. (US) ; Novozymes A/S (DK)

FEATURES

source

1..1208

/organism="Bacillus licheniformis"

/mol_type="unassigned DNA"

/db_xref="taxon:1402"

ORIGIN

Query Match

Best Local Similarity

Matches

50.8%; Score 617.2; DB 6; Length 1208;

70.5%; Pred. No. 88-182;

853; Conservative 0; Mismatches 353; Indels 4; Gaps 2;

QY

1

ATGAGCCAATCGAATTAATGTTTGTAGTGTCTTCTGATCAATTTCAAAACAATCCATAT 60

DB

1

ATGAATCAATCCCTAATAACATTTAGCGTGTGTGAGAACATATACGAAATCCGATAT 60

QY

61

GCTTATTTTCAAACTGCGGAGGAAGATCCGGTTCATTTATGAAGATCGATAGACAT 120

DB

61

CAGTATTTCTCATACCTTCGGGAATCTGATCCGCTCCATTATGAAGATCGCTGGACAG 120

QY

121

TATTTTATCAGCGCTATCATATGTCGCTATATCTTTCAGATCCGATATCTTCACG 180

DB

121

TATTTTATCAGCGCTATCCAGGATGTCGCCGCTGCTCCAGAAATCAGACGCTTTTACA 180

QY

181

ACGAAAT--CACTGTTGAGCGTGGCAACCACTGATGAGGCCCTGTGCTGGCCCAAA 238

DB

181

ACGATTTGCTGCTTGTGCAAAACGGGCGGCCGCTATGCGGACCTGTGCTGCGCCAAA 240

QY 281 GAAGCTTTATCGGTGACGCACTGGATC---ATCTGTCTCATTGATTAACAAAATGCAG 337
Db 314612 AAGGACTTTCAGAGATTATTTTATTCGTACTACGAACCATGATTCGAAGATTACTG 314553
QY 338 AAACTTGTTAGCGCTTATCTTTGAAAGAGGAAAGTATCGTCAATGATTTTGGAA 397
Db 314552 AAGATCTCTTCAACCTTATATGGAAGAAATATAGATATGTTAATGACTTTGGCC 314493
QY 398 AGACCTTTGCGGTGTCGTCACGATGGACATCTCGGGTGGATAAAGAGACCATGAA 457
Db 314492 GCGATTATGCCGTATTAGTGAATTATTCGGAGTGCCTCAAGTGAATACTATCGT 314433
QY 458 AAATCTCTGAGTGGCACAGCGAGTTGCCGATTTTATACAGAGTATCTCTCAATCTCCTG 517
Db 314432 ATATAGCTGAATGGCACAGGGTATCGCCAGTTTATTAACCCAGTTTGACCAACAGAAC 314373
QY 518 AAGCGGGGCAATTCGTTTATGGTGAGGAGGAAAGCTTTCCCAATACATCTGATCGCGTCA 577
Db 314372 TAGAAAAAATGCACAGTCTTGAATGTAGTCAGAAAATTAATTCGCTTACTTAAAGCCTATA 314313
QY 578 TTAAAGAGCGTCGCTCAATCCGGATCAGATTTAAATTTTCGATCCTATGCTACTTCTGAT 637
Db 314312 TAGATCAGAGACGGGTATATCCAGTAAGGACATTAATATCTATATCTG----- 314264
QY 638 ATGAAGGATGCGCTGTGCGACAAGGATATATCTCGCACTGATTTCTTAATGTGCTGTAG 697
Db 314263 -TCAAGATACAGGATGTCATGAGTGAATTAACAGCACTGTGTTAAACATCTTATTAG 314205
QY 698 CGCAACGGAAACCGGCTGATTAAGACGCTGGCACTGATGATCTACCAATTTGCTCAACAATC 757
Db 314204 CGGCTACAGAACTGTGACAAGATATTAGCAATGATGCTTAATCACTTAATCTAATC 314145
QY 758 CTGACGAGATGAATGATGTTTGGCTGACCGTTCGTTAGTTCGGAGAGCCATTCGGAGA 817
Db 314144 CGAGTATGCTTGATGTAGTCTCAAGATTCGAGCTTAGTTCGAGATGCATTTGAAGAA 314085
QY 818 CATTCGGTTTATAAACCGCGGTTTCAGCTGATTCGCGGCGAGCTGTCCTCAAGATACAGTGG 877
Db 314084 CTTTAGCTTGACATCCCGCTACACTCATTTCCAGAGAGCGAGTGAGACGTCACCTA 314025
QY 878 TCGCGGATGAAATCAAAAAGATACGATTTGTTTGTATGATCGGTGCGGCTAACCC 937
Db 314024 TTCTGGTATCGATATACCTAAAGGCGCTGTAGTATTTGTATGATTTGGCGCAGCTAATC 313965
QY 938 GGGACCTGAACCATTTGAACAGCTGACGTGTTTAATATTCTATCGGAGATCTTGGA 997
Db 313964 GTGATCCATCGGTTTTTCAAAACCAAAATGAATTTGATCTATATCGAAGAAAAATACCA 313905
QY 998 TCAAGAGCGCTTTTAGCGCGCGCGCGCATCTCGCTTTCCGATCCGGCATTCATAACT 1057
Db 313904 CTTCTCCACAGAAACCAATCGNAAAGACATTTAGCTTTGGTGCAGGTACTCATGCT 313845
QY 1058 GTGTAGGACAGCTTTTGCNAAAGAAATCGAAATTTAGTACTAATATTGCTGGATA 1117
Db 313844 GTGCTCGGCTGCAATTTTCACTCAGTCACTGTTGAAGTTTCATCCAATATTATTTGGATC 313785
QY 1118 AGATCGGAATACAGATTAGAGGAGATTTTGTATGCTGAGTCCGCTCTGTATACAC 1177
Db 313784 TCCTGATAAATTACGTTTTTGTGATCACTATATATCAAGAAACAGCGGTATACAC 313725
QY 1178 GCGGACCTGTTTCACTTCTCGTTGCGTTGA 1208
Db 313724 GAGGACCTCTAAACTCCTTTGAGTTTGA 313694

RESULT 5
LOCUS AX770907 349980 bp DNA linear PAT 02-JUL-2003
DEFINITION Sequence 38 from Patent WO02094867.
ACCESSION AX770907
VERSION AX770907.1 GI:32438071
KEYWORDS Photorhabdus luminescens

ORGANISM Photorhabdus luminescens
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Photorhabdus.
1
REFERENCE
AUTHORS Duchaud,E., Taourit,S., Glaser,P., Frangeul,L., Kunst,F.,
Danchin,A. and Buchrieser,C.
TITLE Sequence of the Photorhabdus luminescens strain TT01 genome and
uses
JOURNAL Patent: WO 02094867-A 38 28-NOV-2002;
INSTITUT PASTEUR (FR); CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE (CNRS) (FR)
FEATURES
source
1. 349980
/organism="Photorhabdus luminescens"
/mol_type="genomic DNA"
/db_xref="taxon:29488"
/note="sequence length to big . Split into 2
sequences.--seq 00038: 1 too 349980--seq 10787: 300001 to
618776"
ORIGIN
Query Match 20.4%; Score 247.4; DB 6; Length 349980;
Best Local Similarity 52.7%; Pred. No. 4.3e-65;
Matches 617; Conservative 0; Mismatches 536; Indels 18; Gaps 3;
QY 41 AATTTCAAAACAATCCATATGCTTATTTTTCACAACCTCGGGAGGAAGATCCGGTTTCATT 100
Db 283246 AATTCATAAAATCTCTTACGATTTTACGACATTTTGTGATTAACAAGATTTAGTTTATT 283305
QY 101 ATGAAGAGTCGATAGACAGTATTATTTATCAGCGCTATCATGATCTCGCTATATCCTTC 160
Db 283306 TTGAACAATCCAAAATAGTATTTTATTTGCAATATAGGATGTGAGCGTATTTTAA 283365
QY 161 AGCATCCGGATATCTTACGACGAAATCACTTGTGAGCGTCCGACCAACAGTCATCGCGAG 220
Db 283366 AATCCTCA--ATTTCATATACTAAGCCATTAACATGCACTTCCCGAACCGGTATGGGG 283422
QY 221 GCCCTGTCTCGCCCAATGATGGAAGAAGACACTCTGCCAAAAGAGAAATTTGTAGTCA 280
Db 283423 ATCGGTTCTTGCTCAAAATGGAAGTGAAGCATGCGTGTAAACGAAATTCATAATGC 283482
QY 281 GAAGCTTTATCGGTGACGCACCTGGATC---ATCTGTCTCCATTGATTTAAACAAAATGCAG 337
Db 283483 AAGGACTTTCAGAGATTATTTTAAATCGTACTACGAACCCATGATTCGAAAGATTACTG 283542
QY 338 AAACTTGTTAGCGCTTATCTTGAAGAGGAAAGTATCTCTCAATGATTTTGGAA 397
Db 283543 AAGATCTCTTCAACCTTATATGGAAGAAAGGAAATATAGATATTGTTAATGACTTTGGCC 283602
QY 398 AGACGTTTTCGCTGTCGCTCAGATGGACATCTCGGCTGGATAAAGAGAGACCATGAA 457
Db 283603 GCGATTATGCGTATTAGTGACATTAAGTATTTCTCGGACTGCCAAGTATATCTATCGTG 283662
QY 458 AAATCTCTGAGTGGCACAGCGAGTTGCCGATTTTATCAAGAGTATCTCTCAATCTCCTG 517
Db 283663 ATATAGCTGAATGGCACAGGATTCGCCAGTATTCAGTATTTTATACCCAGTTTGACCAACAGAAC 283722
QY 518 AAGCGCGGACATTCGTTTATGTTGTCAGGAAACACCTTTCCCAATACCTTGATGCCGCTCA 577
Db 283723 TAGAAAAAATGCACAGTCTTGAATGTAGTCAGAAATTAATTCGCTTACTTAAAGCCTATA 283782
QY 578 TTAAAGAGCGTCGCTCAATCCGGATCAGATTTTAAATTTTCGATCCTATCTACTTCTGAT 637
Db 283783 TAGATCAGAGACGGGTAAATCCAGTAAGGACATTAATCTATATCTG----- 283831
QY 638 ATGAAGGATGCGCTGTGCGACAAGGATATATCTCGCACTGATTTCTTAATGTGCTGTAG 697
Db 283832 -TCAAGATACAGCGATGTCATGAGTGAATTTACAGCACTGTGTTTAAACATCTTATTAG 283890
QY 698 CGGCAACCGAACCGGCTGATTAAGACGCTGGCACTGATGATCTACCATTTGCTCAACAATC 757
Db 283891 CGGCTACAGAACCTGCTGACCAAGATATTAGCAATGATGCTTAATCACTTAATATCTAATC 283950

```
OY 758 CTGAGCAGATGAATGATGTTTGGCTGACCGTTCGTTAGTTCGGAGAGCCATTGGGAGA 817
DB 283951 CGAGTATGCTGATGATGTTCTCAAGATCGCAGCTTAGTTCGAGATGCAATTGAAGAAA 284010
OY 818 CATTGGGTTATAAACCGCGGTTTCACTGATTCGGCGGAGCTGTCCCAAGATACACAGTGG 877
DB 284011 CTTTAGCTGACATCCCGTACACTCATTCGAAGAGGCGGAGTGAAGACGTCACCTA 284070
OY 878 TCGGCGGTATGAAATCAAAAAGATACGATGTTTTTTTGTATGATCGGTGGCGGTAACC 937
DB 284071 TTTCTGGTATCGATATACCTAAAGGCGCTGTAGTATTTTGTATGATTTGGCGGAGCTAATC 284130
OY 938 GGGACCTCAAGCATTTGAACGCTGACGTGTTTAAATATTCATCGGGAAGATCTTGTA 997
DB 284131 GTGATCCATCGGTTTTTCAAAACCAATGAATTTGATCTATATCAAGAAAAATACCA 284190
OY 998 TCAAGAGCGCTTTTAGCGCGCGCGCATCTCGCTTTTCGGATCCGCGCATTCATAACT 1057
DB 284191 CTTCTCCAGAAAGCAATCGAAAGACATTTAGCTTTTGGTCAGGTACTCATGCT 284250
OY 1058 GTGTAGAGCAGCTTTTGGCAAAAAGCAATCGAAATGGAATTTAGCTAATATTTGCTGGATA 1117
DB 284251 GTGTCGGCTGCAATTTTCACTCAGTCAGTGAAGTTTCATCCCAATATTTTGGATC 284310
OY 1118 AGATCGGATATCAGATTAGAGGAGATTTTGTATGCTGATCGCTGCTGTATACAC 1177
DB 284311 TCTCGATAATTTACGTTTTTGTGATCATTATCAAGAAACAGCGGTGATACAC 284370
OY 1178 GCGGACCTGTTTCACTTCTCGTTCGCTTGA 1208
DB 284371 GAGGACCTTCAACTCCTTTGGATTTGA 284401
```

```
RESULT 6
LOCUS AY260760
DEFINITION Streptomyces hygroscopicus subsp. yingchengensis Shy1 (shy1), Shy2 (shy2), Shy3 (shy3), Shy4 (shy4), Shy5 (shy5), Shy6 (shy6), Shy7 (shy7), Shy8 (shy8), Shy9 (shy9), Shy10 (shy10), Shy11 (shy11), Shy12 (shy12), Shy13 (shy13), Shy14 (shy14), Shy15 (shy15), Shy16 (shy16), Shy17 (shy17), Shy18 (shy18), Hmr19 (hmr19), Shy20 (shy20), Shy21 (shy21), Shy22 (shy22), Shy23 (shy23), Shy24 (shy24), and Shy25 (shy25) genes, complete cds.
ACCESSION AY260760
VERSION 1
KEYWORDS
SOURCE Streptomyces hygroscopicus subsp. yingchengensis
ORGANISM Streptomyces hygroscopicus subsp. yingchengensis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 36688)
AUTHORS Qin,L., Wang,H., Wu,Z., Lu,Y.T., Deng,Z. and Zhao,G.
TITLE Identification and function analysis of a MDR efflux protein-Hmr19, in a sequenced genomic DNA fragment from Streptomyces hygroscopicus yingchengensis 10-22
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 36688)
AUTHORS Qin,L., Lu,Y.T., Zhao,G., Bao,K., Deng,Z., Ren,S. and Jin,W.
TITLE Direct Submission
JOURNAL Submitted (23-MAR-2003) Genome Laboratory, Shanghai Research Center of Biotechnology, CAS, 500 Cao Bao Road, Shanghai 200233, China
FEATURES
source
1..36688
/organism="Streptomyces hygroscopicus subsp. yingchengensis"
/mol_type="genomic DNA"
/isolate="10-22"
/sub_species="yingchengensis"
/db_xref="taxon:228732"
/clone="pHZ1392"
983..2209
/genes="shy1"
983..2209
gene
CDS
```

```
/gene="shy1"
/note="putative P450 protein"
/codon_start=1
/transl_table=11
/product="Shy1"
/protein_id="AAP21648.1"
/db_xref="GI:30267716"
/translation="MSRAHHLPDILSPFAANPYPAYAVMREKEBLIWHTEATQSVIIS
RYEDVERFVKDKAEFTDNNVYNQLEPVHVKITILQSGREHVAVRALVAPAFRGSDLE
KQFLPVIERNRELIADAFRTGSDADVNDYATFPVNVIAIDMLGDKADHAFHGMYT
AVIAPLNLGSDPEVAAGERTVFEAFYMLFVIRERRANPGDDLLSLTCAAEVDGVR
MSDEDIKAFCSLLLAAGGETTDKAGILANLLSHPDOLAAVRADRSLLTCAAEFTLR
YTTPQVIMROSATDEVTGTTIPAGATVTCIIGAANDRERYRDPDRDIFRDDLAT
TSAFAAAGHLAALFALGRHFCVGCALLAKAEVEVLNQLLDAMPDLRLADGHDLEQGVF
TRGPTLPRFTPTVA"
complement(2270..4150)
/gene="shy2"
complement(2270..4150)
/gene="shy2"
/note="putative P450 like ATP/GTP binding protein"
/codon_start=1
/transl_table=11
/product="Shy2"
/protein_id="AAP21649.1"
/db_xref="GI:30267717"
/translation="WGSATSELPSORTPLTAAETCLKIVVGGVGGKTLTVRSVSE
IRPLTEMTQAGQIDETAGVERKTTTVAFDGRISLNDRMWLYLFGAQGRFW
FLWDLRFAGTAVVLDVTRRMECDWYAIIDREHGTFFVAVNVPDGDGKRFSDV
ROALALGHPWMECDARVRASKEVLIADVHLTRALAKESTACSDTTGFPSTDAP
PPGPAHGSAPVLAQYQOTPSQYRLRREHGAAPVLLDGGIPAWLVGPEVCY
VTADHELPAHDSRRWNEHIPPDMPLPYCYQPSVLTGAEHQRRAVITQALEG
VDQELARECQLIAARLISFSGSRAELMSYAHALPARGVLMCGMPAEDATRL
VDDLRLSDAGEDDPVAAYTRVGERIMRLVKEKRRPGPDVTSRMLHPAGLGEEL
VQDLISVIAAAQPTNMI CNTLRLLTDERFVAVNAGRVSVGALNEVLWLDTPQ
NFIGWAVRDTQLGRRHREGCLVLGAANTDQIWPFHAGSGNSAHLFSFSGEH
RCPYAPLLADVMARTAVETLLEHLPDLVLAPEELTWPSIWMRGLTSLPVEFTP
MN"
complement(5031..6773)
/gene="shy3"
complement(5031..6773)
/gene="shy3"
/note="putative sensor kinase"
/codon_start=1
/transl_table=11
/product="Shy3"
/protein_id="AAP21650.1"
/db_xref="GI:30267718"
/translation="MPIPARSQRGAGGAAAPSPALPVILAVTAAGAATAVLAP
DGRVWGTVVVWVICIAVSVRASVIVRRHRAAEQNELELARNWRHQSATSR
LVDVILPDVQVRNGTGVVEALSTVPSDPLQRLKRVATELRSVGRALAESG
LEDVRAQLERTAEILRRVAETLPAAIALLEGSSADTVLAKFPHNVLLRSSAERF
VRELAYSRRSAAQAASAKALSRVQAKTSMADLREMQERHGEVFGDLRLDHSF
SOLGMDRLALLMGSRSVNMKPIVMESILRGAVRIAAYRRVRLHSSRAAISAF
AAEGVMHLLAELMDNAANFSPIDVHVYVEERSAGLVITIEDSLGKMSDAMRRVA
SVSGRTDLASLQGTSLGLVGVRLAVKXISVSRPSRGGTGVVLLPPLVAQRH
EPLTERPQAPAPHSPARPADEPLTLPPEAPARTAPPAPAPAPATSDVTAGGG
ATANGLPAPAPAPAPAEERTRKPAQTPADPRVPGPALDAGSRFGPHRRRAEG
ASADGTDHTVPDA"
complement(9017..9682)
/gene="shy4"
complement(9017..9682)
/gene="shy4"
/note="putative transcriptional regulatory protein"
/codon_start=1
/transl_table=11
/product="Shy4"
/protein_id="AAP21651.1"
/db_xref="GI:30267719"
/translation="MTVAGKETPRERYRAQLFAEIKERAWEQIAAAGATGLSLNIAK
QMMSGPALRYFASRDELITELRDAYSLLADTIRTAASAGADLAVLGRALRAWALD
DPQRYELIYGTVPVGYHAPDDITATASEIMOALLDAAOPTKTVTVDDGHPDTRLEA
HLAQRQWAAHPAPPAALRRALLFWTRLHGILSLELAGHFTGMGLDPAELVDNELRH
LSQ"
```



```
QY 829 AAACGCGCGTTCAGCTGATCCGCGGACAGCTGTCCCAAGATACAGTGTGTCGGCGGTATG 888
Db 1814 ACACGCGCGTTCAGATGATCATGCGGACAGCTGTGCGGACCGACGTCGAGGTCAACCGTGGC 1873
QY 889 GAAATCAAAAAGATACGATCTTTTTCGTATGATCGGTGCGGCTAACCGGACCTGTAA 948
Db 1874 ACATACCGCGCGGCGGACGCTACCTGTCTATCGGCGCGCGGACGAGGATGAGCGG 1933
QY 949 GCATTGGAACACGCTGACGTGTTTAAATATTCATCGGGAAGATCTTGGTATCAAGAGCGCT 1008
Db 1934 CGTACCGCGGACCCGACCGGTTTCGACATCTTCCGCGACGACTCGCCACGACCTCGGCG 1993
QY 1009 TTATGCGGCGCGCGCGGATCTCGTTCGATCGCGATTCATTAACCTGTGTAGGACA 1068
Db 1994 TTCTCGGACGTCGCGGCGACCTCGCTTCGCGCTGCGCGGACCTTCGCGTGGCGCG 2053
QY 1069 GCTTTTGCACAAAACGAAATCGAAATTTAGCTAATATTTGCTGATGAAGATCGCGAAT 1128
Db 2054 CTCCTGGCCACGCGGAGTTCAGGTGGCTTGAACCACTTCTGACGCCATGCCGAC 2113
QY 1129 ATCAGATTAGAGGAAGATTTTGTATGCTGAGTCCGCTCTGTATACACGCGGACCTGTT 1188
Db 2114 CTGCGCTCGCGGACGCGGACGATCTCGTTGAACAGGGGCTCTTACCGCGGCGCGAAG 2173
QY 1189 TCATTCTCTGT 1199
Db 2174 ACGTGGCGGT 2184

RESULT 7
SC0939106 314100 bp DNA linear BCT 11-FEB-2003
LOCUS Streptomyces coelicolor A3(2) complete genome; segment 3/29.
DEFINITION AL939106 AL096837 AL121746 AL132662 AL132856 AL132991
ACCESSION AL133171 AL133424 AL136149 AL136502 AL137165 AL445945 AL449216
AL591823 AL593822 AL545982
VERSION AL939106.1 GI:24413728
KEYWORDS Streptomyces coelicolor A3(2)
SOURCE Streptomyces coelicolor A3(2)
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomyces; Streptomycetaceae; Streptomyces.
REFERENCE 1
AUTHORS Bentley,S.D., Chater,K.F., Cerdeno-Tarraga,A.M., Challis,G.L.,
Thomson,N.R., James,K.D., Harris,D.E., Quail,M.A., Kieser,H.,
Harper,D., Bateman,A., Brown,S., Chandra,G., Chen,C.W., Collins,M.,
Cronin,A., Fraser,A., Goble,A., Hidalgo,J., Hornsby,T., Howarth,S.,
Huang,C.H., Kieser,T., Larke,L., Murphy,L., Oliver,K., O'Neill,S.,
Rabinowitz,E., Rajandream,M.A., Rutherford,K., Rutter,S.,
Seeger,K., Saunders,D., Sharp,S., Squares,R., Squares,S.,
Taylor,K., Warren,T., Wietzorek,A., Woodward,J., Barrell,B.G.,
Parkhill,J. and Hopwood,D.A.
Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2)
JOURNAL Nature 417 (6885), 141-147 (2002)
MEDLINE 21996410
PUBMED 1200953
REFERENCE 2 (bases 1 to 314100)
AUTHORS Bentley,S.D.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces
sequencing team, Sanger Institute, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk
COMMENT On or before Oct 26, 2002 this sequence version replaced
GI:20520920, GI:20520898, GI:20520901, GI:20520903, GI:20520904,
GI:20520905, GI:20520906, GI:20520907, GI:20520926, GI:20520909,
GI:20520910, GI:20520805, GI:20520706, GI:20520789, GI:20520843.
Location/Qualifiers
1. 314100
/organism="Streptomyces coelicolor A3(2)"
/mol_type="genomic DNA"
/strain="A3(2)"
/db_xref="taxon:100226"

FEATURES
source
```

```
RBS
gene
CDS

53. 56
67. 945
/gene="SCO0550"
/note="synonym: SCF11.30"
67. 945
/gene="SCO0550"
/note="SCF11.30, possible membrane protein, len: 292 aa;
similar to TR:Q9ZBQ4 (EMBL:AL034447) Streptomyces
coelicolor putative transmembrane protein, 396 aa; fasta
scores: opt: 677 z-score: 505.6 E(): 8.6e-21; 43.7%
identity in 309 aa overlap. Contains possible hydrophobic
membrane spanning region"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB59607.1"
/db_xref="GI:6137052"
/db_xref="GOA:O9RK60"
/db_xref="TrEMBL:Q9RK60"
/translation="MGRRAARPSRSHARPRGGLSRGAALALGAVTLTAGVTILAV
GDGDDGASDRVTANAADOAGNAEAGAAPLASASPSATAGSRSPSPAPKRTP
TKTASPATSPKASAPASAPVATKAGDRSGGSGSGSGSGSGSGSGSGSGSG
SGSGTGAAGAOVLALVNEERAAAGCSPVTANSRLTRAADDYSDVMASGVMGHTGPDG
STMASRVAAGYQNSALGENIAQQAADAASVMEAWMNSPGHRENILNCSFKELGVGVH
FGDGPWTQDFGTGR"
complement(1031..2359)
/gene="SCO0551"
/note="synonym: SCF11.31c"
complement(1031..2359)
/gene="SCO0551"
/note="SCF11.31c, possible histidine kinase protein, len:
442 aa; similar to TR:O50497 (EMBL:AL020958) Streptomyces
coelicolor two-component system sensor, 481 aa; fasta
scores: opt: 401 z-score: 451.3 E(): 9e-18; 30.2% identity
in 367 aa overlap. Contains Pfam matches to entries
PF00512 signal, histidine kinase and PF00672 DUF5, domain
found in bacterial signal proteins. Contains possible
N-terminal signal peptide sequence"
/codon_start=1
/transl_table=11
/product="putative histidine kinase protein"
/protein_id="CAB59608.1"
/db_xref="GI:6137053"
/db_xref="GOA:O9RK59"
/db_xref="TrEMBL:Q9RK59"
/translation="NRRALAGIALAVTSMVALSVLPLVLRREQARDVTTAAEQRA
AALSPVALTTRADVQQAELGSADOLVRLPDGFGVTGPHAPPEALDRAVRGRET
LAVDTADGWAYLQPVVLGRDVAVEAYVPAADLRGTASGVMALLAVGLVGSVL
VADRIGARVVRSSRGKRLASLALGSGDLGVRVPEGPPELQAGAAFTMADRVLL
AVERELVADLSHRLTPTLALYLEADRMGATSAKRVTEAGQLRELSIIIAAKTP
LAAAGLGTGATGATKPCDVAEVVAMRLDFWSVLAQAQDPFERSLTPRPAPVAFPEDD
LAAVVDALIGNVFRHTPQCTRFVRVRRDRHRLTVDGAPGVADPEALTRGVSVG
GSTGLGLDIVARAARTADGELTITRAPLGGARVTVSFALADARAAGDGR"
complement(1070..1672)
/gene="SCO0551"
/note="Pfam match to entry PF00512 signal, Histidine
kinase, score 39.60, E-value 7.1e-08"
complement(1712..1921)
/gene="SCO0551"
/note="Pfam match to entry PF00672 DUF5, Domain found in
bacterial signal proteins, score 36.90, E-value 4.5e-07"
complement(2356..3072)
/gene="SCO0552"
/note="synonym: SCF11.32c"
complement(2356..3072)
/gene="SCO0552"
/note="SCF11.32c, possible response regulator, len: 238
aa; similar to TR:CAB54568 (EMBL:AJ006392) Streptococcus
pneumoniae response regulator Rro2, 234 aa; fasta
scores: opt: 583 z-score: 693.9 E(): 2.8e-31; 43.2% identity in
227 aa overlap. Contains Pfam matches to entries PF00486
trans_reg_C, transcriptional regulatory protein and
PF00072 response_reg, response regulator receiver domain"
misc_feature
misc_feature
gene
CDS
```



```
/codon_start=1
/transl_table=11
/product="putative response regulator"
/protein_id="CAB59609.1"
/db_xref="GI:6137054"
/db_xref="GOA:Q9RK58"
/db_xref="TrEMBL:O9RK58"
/translation="WTRLTCLPMASVLVVEDPVIKALIEVLTHGHVAVKTAHQGF
EALRDTQSPDIVLGLDLGLDLVLRMRIGSRVPLVATARDDETEILRLINA
GADDZWRFDSGGQAAARLAARVLRSPAGLAARPPVLRVADLRIDPTARTAHLAGR
ELPLTRREFDLAYLAARADQVMSQRILAEYMQQPYVEDQTVVHLASLRKMGKA
RKPRYLHTVRGIGIKLVAP"
complement(2380..2595)
/gene="SCO0552"
/notes="Pfam match to entry PF00486 trans_reg_C,
transcriptional regulatory protein, C terminal, score
93.00, E-value 1.7e-26"
complement(2710..3042)
/gene="SCO0552"
/notes="Pfam match to entry PF00072 response_reg, Response
regulator receiver domain, score 123.00, E-value 5.6e-33"
complement(3090..3093)
complement(3099..3536)
/gene="SCO0553"
/notes="synonym: SCF11.33c"
complement(3099..3536)
/gene="SCO0553"
/notes="SCF11.33c, unknown, len: 145 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SCF11.33c"
/protein_id="CAB59610.1"
/db_xref="GI:6137055"
/db_xref="TrEMBL:O9RK57"
/translation="MKPFGTAIRHTSPNRTVALPPSEGVTVTPAPSSHPGAPA
PAGRGAVVRGLGVGVTVASGVSGLVRGLVGLVGLVGLVGLVGLVGLVGLVGLV
GAGSFGAVCATGVRVDPPTTKIWMVTAVTLAAVQDSQMR"
complement(3691..5388)
/gene="SCO0554"
/notes="synonyms: manA, SCF11.34c"
complement(3691..5388)
/gene="SCO0554"
/notes="SCF73.01c, manA, probable secreted
beta-mannosidase, len: 327 aa; similar to many e.g.
MANA_STRLI P51529 mannan endo-1,4-beta-mannosidase
precursor (EC 3.2.1.78) (363 aa), fasta scores; opt: 1189
z-score: 1332.7 E(): 0, 61.1% identity in 288 aa overlap.
Contains N-terminal signal sequence, and Pfam match to
entry PF00150 cellulase, Cellulase (glycosyl hydrolase
family 5). Also similar to SC2H4.16 (62.2% identity in 288
aa overlap)
SCF11.34c, possible sugar hydrolase, partial CDS, len:
>278 aa; C-terminus similar to SW:GUN4_THIEFU (EMBL:L20093)
Thermomonospora fusca endoglucanase E-4 precursor (EC
3.2.1.4) (endo-1,4-beta-glucanase E-4), Celd, 880 aa;
fasta scores: opt: 544 z-score: 555.0 E(): 1.5e-23; 42.8%
identity in 215 aa overlap. Contains Pfam matches to
entries PF00553 CBD_2, cellulose binding domain and
PF00041 fn3, fibronectin type III domain. Overlaps with
Streptomyces coelicolor Stf73.01c, len: >327 aa;
C-terminus similar to many e.g. MANA_STRLI P51529 mannan
endo-1,4-beta-mannosidase precursor (EC 3.2.1.78) (363
aa), fasta scores; opt: 1189 z-score: 1332.7 E(): 0, 61.1%
identity in 288 aa overlap. Contains N-terminal signal
sequence, and Pfam match to entry PF00150 cellulase,
Cellulase (glycosyl hydrolase family 5)"
/codon_start=1
/transl_table=11
/product="putative secreted beta-mannosidase"
/protein_id="CAD55266.1"
/db_xref="GI:24413729"
/db_xref="GOA:Q8CK53"
/db_xref="TrEMBL:Q8CK53"
```

```
translation="MRPARRDTPRTLPARLLTGLAALIGLVVVVVGALCPGAALAQSPAA
RSAGVPSAGAAAGLHIGDGLLEGNDKRVNMGVNHATHTYPGTQSLADYKALGAN
SVRVLSGCHRWSENGPADVAAVIEQCKXANRLICVLEVHDTTGYAAAGLQHTIMVDAP
YWIIGLVLAGOEDYIVNIENBPMTDTPAGTPTVAAVKKRAAGLQHTIMVDAP
NWGDQGVMRANARSVYDADPTGNLIFSIHMYSVDTAQEITDYINAFVDAELPILI
GEFGPDQYGDGPDDEMTMAEAQLGLYLAWSWGNTPDVLDDALDFPDLRSLWGE
RI FHVGNVIGIAOTSEBATVYGDGAPDTPBAPGAPSVTAVTDSVSTLGMVAATDDTA
VSGVDVLVGGDTSSTVASSTTTATATVGLTAATYTRAVHARDAGNPSASATVEA
TIDGGTTPGGACSGVGRVGVGWPFGFQSEIAVRNTGSAAGIHPWTILAFAPADQQTVMN
WGTATDQAGAVSVAPASYSYATIPAGGTIVLGTFTARKGDTNTAPAFRLUNGAVCATP"
complement(3703..4005)
/gene="SCO0554"

misc_feature
Query Match
Best Local Similarity 13.9%; Score 169; DB 1; Length 314100;
Matches 445; Conservative 0; Mismatches 460; Indels 0; Gaps 0;

QY 295 GACGCACTGGATCATCTGTCTCCATTGATTAAACAAAATTCAGAAAACTTTGTAGCGCCT 354
|||||
Db 35474 GACCTCCAGAGCGGTTCTTCGCGGTTCATCGAGCGCACTCCCGTGAGCTGATCGACGG 35533
QY 355 TATCTTTGAAAGAGGAAAAGTGAATCTCGTCAATGATTTTGGAAAGACGTTTGGGTGTGC 414
|||||
Db 35534 TTCCGGCACACCGCGCGCGCGATCTGGTCGCGCACTACGCGACCGGTTCCCGGTCAAC 35593
QY 415 GTCACGATGGACATCTCGGCTCGGATAAAGAGACCATGAAATAATCTCTGAGTGGCAC 474
|||||
Db 35594 GTCATCGCGACATCTGGGCTCGGACCAAGGCGCACGACCGGTTCCACGGGTGGTAC 35653
QY 475 AGCGGAGTTGCGGATTTTATCACGAGTATCTCTCAATCTCTCTGAAAGCGGGCACATTCG 534
|||||
Db 35654 AGTGCGTTCATCGCTTCTCGGCAACCTCTCCGGGACAGGAGTCCGCGGGCGCGGT 35713
QY 535 TTATGGTTCAGCGAAACAGCTTTCCCAATCTTGTATGTCGCGGTCAATTAAGAAACGTGCGCTC 594
|||||
Db 35714 GCGCGCACCGGACGGAGTTCGCGAGTACATGATCCCGGTTCATCGGAGCGTCTGTGAG 35773
QY 595 ATTCGGGATCAGATTTAATTTTCGATCTATGATCTCTGATATATGAAGGCATGGCGTG 654
|||||
Db 35774 AACCGGCGGACGACCTGCTGTCCACGCTGTGCGCGCGGAGGTGGACGGCTCGGATG 35833
QY 655 TCGGACAAAGGATATCTCGCACTGATTTCTTAATGTGCTTTAGCCGCAACGAAACCGGCT 714
|||||
Db 35834 AGCGACGAGGACATCAAGCGGTTCTGCGAGTCTGCTCGCGCGGGCGGCGAGACCAC 35893
QY 715 GATAAGACGTGGCACTGATGATCTACCATTTGTCTCAACAATCTCTGAGCAGATGAATGAT 774
|||||
Db 35894 GACAAGGCGATCGCGCGCATCTTCGCCAACCTGCTGGGCCCATCCGAGCAGTTGGCGGCC 35953
QY 775 GTTTTGGCTGACCGTTCTGTAGTTCCGAGAGCCATTGCGGAGACATTCGGTTATTAACCG 834
|||||
Db 35954 GTGCGGAGGACCGGAGCGCTGATCCCGCGCGCTTCGCGGAGACGCTGGCTGACACCGC 36013
QY 835 CCGGTTACAGCTGATTTCCCGCGGAGCTGTCCCAAGATACAGTGGTGGCGGCTATGAAATC 894
|||||
Db 36014 CCGGTTCCATGATCATCGGTGAGCGGACCGGACGCTACGCTCAGCGCGGCGCAGATA 36073
QY 895 AAAAAAGATACATTTGTTTTTTGATGATCGGTTCGGCTTAACCGGACCCCTGAAAGCATTT 954
|||||
Db 36074 CCGGCGGTGCGCACCGTCCACTGCTGATAGGAGGGGCCAACCGGACGAGACGGGCTAC 36133
QY 955 GNACAGCTGACGTTTAAATATTCATCGGAGATCTTGGTATCAAGAGCGCTTTTACG 1014
|||||
Db 36134 CCGGACCGGACCGCTTCGACATCATGCGGACGACCTGACACGACGACCGCTTCTCG 36193
QY 1015 GCGCGCGCGCGCATCTCGCTTTTCGGATCCGCGATTCATTAATCTGTGTAGGACGAGCTTTT 1074
|||||
Db 36194 GCGCGCGCGACCTTCGCTTCGCGTTCGCGCGGACCTTCGCTGCGCGCGCTGCTG 36253
QY 1075 GCCAAAAACGAAATCGAAATTTGTAGCTAATATTTGTGTGGATAGATCGGAAATATCAGA 1134
|||||
Db 36254 GCGAAGGCGGAGGTGAGATCGGGTTCGCTCAGCTCTCTGGACGCGCTGCCCGCGCTCGCG 36313
```


QY 1135 TTAGAGGAAGATTTTGTATGTCGAGTCGCGTCTGTATACACGCGGACCTGTTTCACCTT 1194
 |||||
 Db 36314 ACGGAGGACGGCTTCGAGTCGTGCGAGCGGCGCTGTTCCACCGCGCGCGCAGTCGCTG 36373
 |||||
 QY 1195 CTCGT 1199
 |||||
 Db 36374 CCGGT 36378

RESULT 8
 AEO12559
 LOCUS AEO12559 AE009442
 DEFINITION Xylella fastidiosa Temeculal, section 7 of 9 of the complete genome.
 AEO12559.1 GI:28057550
 SOURCE Xylella fastidiosa Temeculal
 ORGANISM Xylella fastidiosa Temeculal
 Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales; Xanthomonadaceae; Xylella.

REFERENCE 1 (bases 1 to 300885)
 AUTHORS Van Sluys, M.A., de Oliveira, M.C., Monteiro-Vitorello, C.B., Miyaki, C.Y., Furlan, L.R., Camargo, L.E.A., da Silva, A.C.R., Moon, D.H., Takita, M.A., Lemos, E.G.M., Machado, M.A., Ferro, M.I.T., da Silva, F.R., Goldman, M.H.S., Goldman, G.H., Lemos, M.V.F., El-Dorry, H., Tsai, S.M., Carrer, H., Carraro, D.M., de Oliveira, R.C., Nunes, L.R., Siqueira, W.J., Coutinho, L.L., Kimura, E.T., Ferro, E.S., Harakava, R., Kuramae, E.E., Marino, C.L., Gigliotti, E., Abreu, I.L., Alves, L.M.C., do Amaral, A.M., Baia, G.S., Blanco, S.R., Brito, M.S., Cannavan, F.S., Celestino, A.V., da Cunha, A.F., Fenille, R.C., Ferro, J.A., Formighieri, E.F., Kishi, L.T., Leoni, S.G., Oliveira, A.R., Rosa Jr., V.E., Sasaki, F.T., Sena, J.A.D., de Souza, A.A., Truffi, D., Tsukumo, F., Yanai, G.M., Zaros, L.G., Civerolo, E.L., Simpson, A.J.G., Almeida Jr., N.F., Setubal, J.C. and Kitajima, J.P.
 TITLE Comparative Analyses of the Complete Genome Sequences of Pierce's Disease and Citrus Variegated Chlorosis Strains of Xylella fastidiosa
 J. Bacteriol. 185 (3), 1018-1026 (2003)

JOURNAL 12533478
 PUBMED
 REFERENCE 2 (bases 1 to 300885)
 AUTHORS Van Sluys, M.A., de Oliveira, M.C., Monteiro-Vitorello, C.B., Miyaki, C.Y., Furlan, L.R., Camargo, L.E.A., da Silva, A.C.R., Moon, D.H., Takita, M.A., Lemos, E.G.M., Machado, M.A., Ferro, M.I.T., da Silva, F.R., Goldman, M.H.S., Goldman, G.H., Lemos, M.V.F., El-Dorry, H., Tsai, S.M., Carrer, H., Carraro, D.M., de Oliveira, R.C., Nunes, L.R., Siqueira, W.J., Coutinho, L.L., Kimura, E.T., Ferro, E.S., Harakava, R., Kuramae, E.E., Marino, C.L., Gigliotti, E., Abreu, I.L., Alves, L.M.C., do Amaral, A.M., Baia, G.S., Blanco, S.R., Brito, M.S., Cannavan, F.S., Celestino, A.V., da Cunha, A.F., Fenille, R.C., Ferro, J.A., Formighieri, E.F., Kishi, L.T., Leoni, S.G., Oliveira, A.R., Rosa Jr., V.E., Sasaki, F.T., Sena, J.A.D., de Souza, A.A., Truffi, D., Tsukumo, F., Yanai, G.M., Zaros, L.G., Civerolo, E.L., Simpson, A.J.G., Almeida Jr., N.F., Setubal, J.C. and Kitajima, J.P.
 TITLE Direct Submission
 JOURNAL Submitted (17-DEC-2001) Instituto de Biociencias, Universidade de Sao Paulo, Rua do Matão, 277, Sao Paulo, SP 05508-900, Brazil

FEATURES
 source
 1. 300885
 /organism="Xylella fastidiosa Temeculal"
 /mol_type="genomic DNA"
 /strain="Temeculal"
 /db_xref="taxon:183190"
 /note="Pierce's disease strain"
 87..1022
 /gene="apbe"
 /locus_tag="PD1557"
 87..1022
 /gene="apbe"
 /locus_tag="PD1557"
 /note="identified by sequence similarity; putative; ORF"

located using Blastx/Glimmer"
 /codon_start=1
 /transl_table=11
 /product="thiamine biosynthesis lipoprotein Apbe
 precursor"
 /protein_id="AAO29399.1"
 /db_xref="GI:28057551"
 /translation="MGTTWSVKLAAPRRLDLHLPHNAIOARLDQIWAEMSTWEPNSHI
 SRFNAPVRSWHPLFNDPDLVRLTALDIAMTSNGAPDPTVGLVWLGFGAYSGNHKQ
 IPTPEAITLATRVGWQHLERTTDCHTWLPQGGTDLDSIAIKGVSDVAFAATLHAKGV
 HHALVDIGELGYGKHKPDGTPWSLVEMDLTHOYNPLPPCIIQLDGLAVATSGDRW
 HHFEGORYTHTIDPHQGTPIPOAPALVTMIASSAHADAWATALNTILGREAGLALA
 NTGLAVRYLEHNDAFTAYSPAFTRLLHHPATA"
 complement (1752..2081)
 /gene="come"
 /locus_tag="PD1558"
 complement (1752..2081)
 /gene="come"
 /locus_tag="PD1558"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer"
 /codon_start=1
 /transl_table=11
 /product="DNA transport competence protein"
 /protein_id="AAO29400.1"
 /db_xref="GI:28057552"
 /translation="MKSFAMRLKSLLLFLVLAFTVHAANKVDINTASAEEMDKVLVNI
 GPSKASATVYKREENGPFKSVBELALVKIGMKTVERNRLIEIGTRMAPAKHAKGTE
 LKSVGKR"
 2603..3625
 /locus_tag="PD1559"
 2603..3625
 /locus_tag="PD1559"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAO29401.1"
 /db_xref="GI:28057553"
 /translation="MTSSDFSDTTRATORLOWIRTTADSTATLORASSDAGFRSVWR
 STSNGLSHLNDAPHLNPNQWLRIRHALLTSGGVVRPHILAOQLEAGFLLEDGLIP
 TLAQRDLADNADTLFDAALDQIALQICVPPNDLPRFTALLERDASLFEDFLRHL
 NLTLNRDLALQVQQQIMNALTQPRVLVHRDFMRLMLTTDGVTLVDFDQCTVG
 PVAYDPVSLFKDTSVSWPLRVLRWLYHARANAAPVQTLPHFLRDADWMVQQRH
 LKNLGIIFARLHYRDKGSWYLENIPRFISYLEILLPHRPTLAPLASLIEHRIKPALAAR
 MITEST"
 3622..4344
 /locus_tag="PD1560"
 3622..4344
 /locus_tag="PD1560"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer"
 /codon_start=1
 /transl_table=11
 /product="nucleotidyl transferase"
 /protein_id="AAO29402.1"
 /db_xref="GI:28057554"
 /translation="WKALIFAAGIQRMRLTNYTPKLLCAGGPELIVNLRKLAAL
 GISEVVINTAWLSQFPBILGDGQRFGLRFLYSNEGSLPLETGGGMLHALPILGNAPP
 LAINGDIWTDADLRLTPTEPVGDADHLMLVNNPPEYHPQDGFVLQADSSVLDRTPGIPL
 TFAGLIYRSQLADWRNIIGDTPDTHAQPPRFKPLAPLLRAAMRSRIRHGHGQWT
 DVGTPQRLHALDTWLRSPEARF"
 complement (4527..5228)
 /locus_tag="PD1561"
 complement (4527..5228)
 /locus_tag="PD1561"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer"
 /codon_start=1
 /transl_table=11
 /product="replication related protein"
 /protein_id="AAO29403.1"

gene	/db_xref="GI:28057555"	
	/translation="MSVSQIPLALRYSSDQRFETYLHAPSGLIAQLRALVDQCSREWI YLVGSGKTHLALCAAQVQSVAYLPLOVAVGRLRLDALEGCVLVLDGL QAIWGREDELTFDFNRAAGITLLYARAI PDGLGLTLPDLSRVAQCIRALP TLDDVGRASVLRERAQRGLMIDEAAIEWLLSRGRDRDLGSLVMLLERLDRESLATKR VTIPFLRHVLAEGGG"	
	complement (5225.. 6382)	
	/gene="perM"	
	/locus_tag="PD1562"	
	complement (5225.. 6382)	
	/gene="perM"	
	/locus_tag="PD1562"	
	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"	
	/codon_start=1 /transl_table=11 /product="permease" /protein_id="AAO29404.1" /db_xref="GI:28057556"	
CDS	/translation="MMFSPVEIARPLRLKWWVLVGGVLMWMLLTPVLIPVLSVF LGNLADPLKRVGVGRAMAVCFVFLVLLLVLLMIFVPLIERQVTLINTEFPQ MHDVNTVTPWEQTSQRLRWLDPEQMIQMHNSNEQAGVGARGLFGYVSRGFV MYTWNTLALLIFYLDRWDKWRVAAAIPRAYIGTVSRLEANEVLGAFIRG QFLVMLGAIYAVGSLVGLRGLLIGLIAGLISFIPYLGAITGVWLALIAVIVQH GLDQLLVGVGVGGLLESYVLTPIRVGDKILHPVAVIFSVMAGQLFGFVGM LALPYAAVNVLLRYAHERYIESEFYKGYQLAVALHECKSDENCDGKKDAV"	
	complement (6379.. 7551)	
	/locus_tag="PD1563"	
	complement (6379.. 7551)	
	/locus_tag="PD1563"	
	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"	
	/codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAO29405.1" /db_xref="GI:28057557"	
	/translation="WAMPYFCQQSLFCISHWELSMRYRPFELTALCAVLVPVAVA QRPHIQSDVAKEQSIYEAEPVNSQGSDDRKAGIARALISVLGVTGDRSNAKLRP EVSQALRDAADVSDYSDYDQGSFGGAPTSRTMLTVHFRDEVNALVALPLWP QPRKPVLMALIDDSGRLVGVQSRASVYLDRAIERGRLGLPAGRSBEQALVTA IWRDVSALVSTRYTSMLQVGLKYSRSGNGMAADWIFVGGEOVSSSSNADALR VNADGAADALVKRYAKVPLTGPVTVRVGIRGIRGADDLVRYSAALQRPVVRSM IPVKASABRLEVYLDLMTGITGLNRLMGDSALQPVAAATLSPQSEGLVEYVLK"	
	7602.. 8657	
	/gene="purM"	
CDS	/locus_tag="PD1564"	
	/note="synonym: purG"	
	7602.. 8657	
	/gene="purM"	
	/locus_tag="PD1564"	
	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"	
	/codon_start=1 /transl_table=11 /product="5'-phosphoribosyl-5-aminoimidazole synthetase" /protein_id="AAO29406.1" /db_xref="GI:28057558"	
	/translation="MTTPPPSTPLTYRDAGVIDAGNALVERIKPLVKTFRPEVM SGLGFGFGLNLAGKPELVISGTDVGTGKLKLAQLNRHNTIGIDLVAMCVNDVLV QGAELPFLDYPATKLDIDTATAVISGALCEQSGCALIGETAEMDMYPPGEYD LAGFCVAAVEKSQLDGSQVREDDVLIGIASGPHSNGYSLIRRIYERAGSPADLDIH GRLDITLMAPTALYKPLKULHTSDRAIHAHAHTTGGGLTENIRVPPNLGURD ANAWQPPVFOVQLQREGALADMTWRTFNCIGFGVLVATPNQVAPLQALDNQGLAH QIGRVVTPVDNERVHIG"	
	Query Match 9.6%; Score 116.2; DB 1; Length 300885; Best Local Similarity 49.4%; Pred. No. 6.8e-24; Matches 301; Conservative 0; Mismatches 308; Indels 0; Gaps 0;	
	QY	375
DB	156009	TGATCTGTGCTGAATTTGCCTTTCCATGCGAGATCATTTGCAAAATGATGGA 156068

QY	435	GCTGGATAAAGAGACCATGAAAAATCTCTGAGTGGCACAGCGAGTTGCCGATTTTAT	494
DB	156069		
DB	156069	TGTGGATATTGGAGATGCCGTCACCTCGGTATGCGGTGAGCAAAATTCGCAAGGTACT	156128
QY	495	CACGAGTAGTCTCTCAATCTCTGAAGCGGGGCACATTCGTTATGGTGCAGCGAACAGCT	554
DB	156129	CGACCTTCGCCAATGTCGGCCGATGAATTAGTACACGAAGCACTGCCTACGAAGAGCT	156188
QY	555	TTCCCAATACTTGATGCCGGTCAATTAAGAAACGTGCGTCAATCCGGGATCAGATTTAAT	614
DB	156189	TGCGCAATACTTTACGAAGTTGATTGAGCTACGCCCGCACCCATCTCTGGAACCTGACTTAAT	156248
QY	615	TTGCAATCCTATGTAATCTCTGAATATGAAGCATGGCGCTGTCGGACAAGGATATACTCGC	674
DB	156249	TTCCATGTTTCTGCGGTCTGAGGAAGATGGCGAGAACTAACCCATGATGAATTCGTCTC	156308
QY	675	ACTGATCTTAAATGTGCTGTTAGCCGCAACCGGAACCGGCTGATAAGACGCTGGCACTGAT	734
DB	156309	CAATGTGATTATGTTATTAAATCGCCGTTACGAACCCACATCCAATATGATTGGCAATGC	156368
QY	735	GATCTACCAATTTGCTCAACAATCTCTGAGCAGATGAATGATGTTTGGCTGACCGTTCGTT	794
DB	156369	ATTGATTGCGTTGCAATCGGCATCCAGAGCAACTCGCACTGCTCAAGAGTGATCTGTCACT	156428
QY	795	AGTTCGGAGAGCCATTGCGGAGACATTTGGGTTATAAACCCCGGTTTCAGCTGATTCCGCG	854
DB	156429	GATGCCACAAGCGGTATCGAATGCTTACGTTACGACGGTTCGGTACAGTTCAGATGCG	156488
QY	855	GCAGCTGTCCCAGATACAGTGGTTCGGCGGTATGGAATCAAAAAGATACGATTTGTTTT	914
DB	156489	CGCGCGATGGATGATATAGAGGTTGAAGCGAGTTGGTCCCTCGCGCACAGTAGTATT	156548
QY	915	TTGTATGATTCGGTTCGGCTTAACCGGACCTCTGAAGCATTTGAACAGCTGACGGTTTAA	974
DB	156549	TTTGATGCTTGGTGTCTGCCAACCGTGATCCGCGACAATTCATCTATCCCGACAGTGGA	156608
QY	975	TATTATCTCG 983	
DB	156609	TATTACTCG 156617	
RESULT 9			
AE003889/c			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Xylella fastidiosa 9a5c			
Xylella fastidiosa 9a5c			
Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;			
Xanthomonadaceae; Xylella.			
REFERENCE			
1 (bases 1 to 10092)			
AUTHORS			
Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M., Alvarenga,R., Alves,L.M., Araya,J.E., Baia,G.S., Baptista,C.S., Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R., Bueno,M.R., Camargo,A.A., Carraro,L.E., Carraro,D.M., Carrer,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C., Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorfy,H., Faciniani,A.P., Ferreira,A.J., Ferreira,V.C., Ferto,J.A., Fraga,J.S., Franca,S.C., Franco,M.C., Frohme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A., Ho,P.L., Hohnel,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P. and Marino,C.L.			
TITLE			
The genome sequence of the plant pathogen Xylella fastidiosa. The			
Xylella fastidiosa Consortium of the Organization for Nucleotide			
Sequencing and Analysis			
NATURE 406 (6792), 151-157 (2000)			
JOURNAL			
MEDLINE			
PUBMED			
10910347			
REFERENCE			
2 (bases 1 to 10092)			
AUTHORS			
Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,			

Alvarenga, R., Alves, L. M. C., Araya, J. E., Baia, G. S., Baptista, C. S., Barros, M. H., Bonaccorsi, E. D., Bordin, S., Bove, J. M., Briones, M. R. S., Bueno, M. R. P., Camargo, A. A., Camargo, L. E. A., Carraro, D. M., Carter, H., Colaço, N. B., Colombo, C., Costa, F. F., Costa, M. C. R., Costa-Neto, C. M., Coutinho, L. L., Cristofani, M., Dias-Neto, E., Docena, C., El-Dorri, H., Pacinciani, A. P., Ferreira, A. J. S., Ferreira, V. C. A., Ferro, J. A., Fraga, J. S., Franca, S. C., Franco, M. C., Frohme, M., Furlan, L. R., Garnier, M., Goldman, G. H., Goldman, M. H. S., Gomes, S. L., Gruber, A., Ho, P. L., Hoheisel, J. D., Junqueira, M. L., Kemper, E. L., Kitajima, J. P., Krieger, J. E., Kuranae, E. E., Laigret, F., Lambais, M. R., Leite, L. C. C., Lemos, E. G. M., Lemos, M. V. F., Lopes, S. A., Lopes, C. R., Machado, J. A., Machado, M. A., Madeira, A. M. B. N., Madeira, H. M. F., Marino, C. L., Marques, M. V., Martins, E. A. L., Martins, E. M. F., Matsukuma, A. Y., Menck, C. F. M., Miracca, E. C., Miyaki, C. Y., Monteiro-Vitorello, C. B., Moon, D. H., Nagai, M. A., Nascimento, A. L. O., Netto, L. E. S., Nhani Jr., A., Nobrega, F. G., Nunes, L. R., Oliveira, M. A., de Oliveira, M. C., de Oliveira, R. C., Palmieri, D. A., Paris, A., Peixoto, B. R., Pereira, G. A. G., Pereira Jr., H. A., Pesquero, J. B., Quaggio, R. B., Roberto, P. G., Rodrigues, V., de M. Rosa, A. J., de Rosa Jr., V. E., de Sa, R. G., Santelli, R. V., Sawasaki, H. E., da Silva, A. C. R., da Silva, F. R., da Silva, A. M., Silva Jr., W. A., da Silveira, J. F., Silvestri, M. L. Z., Siqueira, W. J., de Souza, A. A., de Souza, A. P., Terenzi, M. F., Truffi, D., Tsai, S. M., Tshako, M. H., Vallada, H., Van Sluys, M. A., Verjovski-Almeida, S., Vettore, A. L., Zago, M. A., Zatz, M., Meidanis, J. and Setubal, J. C.

Direct Submission
Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP 13083-970, Brazil

Location/Qualifiers
1. 10092
/organism="Xylella fastidiosa 9a5c"
/mol_type="genomic DNA"
/db_xref="taxon:160492"
/clone="9a5c"
715. 1482
/locus_tag="XF0374"
715. 1482
/locus_tag="XF0374"
/locus_tag="XF0374"
/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF83184.1"
/db_xref="GI:9105206"
/translation="MTYNSILSIRINQACIQARKFLPFLHLPLVGVMLPFLGLMG
LAGDQWADLLRLSGGQWLLKQHWITEYLHVRGKWLIVAGLCLGLMLVRCRGLG
AGCYRLASLYVVVTLVSTLSILVKRLTGMDPCWDLVRYGGPPFGLPESRHLK
ASGCFPAGHASAGYAVWCLYFAAWGVCFAWRWGLMGLMIVAGLIFGISQOLRGHFLS
HDLWSLTICWLVAGLGFYLFVSPVTRVVSAQQRANT"
1479. 3137
/locus_tag="XF0375"
1479. 3137
/locus_tag="XF0375"
/note="similar to SP130845 (percent identity: 36 %/query alignment coverage: 98.0 %/subject alignment coverage: 98.9 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="inner membrane protein"
/protein_id="AAF83185.1"
/db_xref="GI:9105207"
/translation="MSFSTLVILVRRMSEFNRVRPVESTSVLGTSLFALACNT
MFWRSAMTSAGSISIGVLSALLVTYHALLGLVWVRWAKPLTLFLVITAFATHY
MNSYSYLDADLRNVFNDHRESLELTLSALIPLLFYAGPIAVLWLRFRQRPWS
RALGRMLLVIVVVGAGCMLSFOLSKALNDRVRLATPINYIMALKMSLNSD
SVLKPSKPIGEDAVATPPIPSRPLIAVIVGETARAQNLGNGVROTPPOLAON
DVINPDMHSCGTNTEVSPCMFSPYGRNRYDESKIRHQSLHLVLERISTLWRDN
QSGCKGVCLQQLQDNDKADPTLCSSGRCDMEILLKDFVQVRSKLGDVRVVVHQL
GSHGPSYFORYFVAFQFSPPTCNLGSCEQVVAAYDNSLLTYDHLFVLTIGLRL
DMSDYDTAMITVLSHGESIGSLGKGLYHGMFYAIPAEQTRVPMVWFQSKQFVQSRQID

TITLE JOURNAL

FEATURES source

gene

CDS

/note="similar to SP130845 (percent identity: 36 %/query alignment coverage: 98.0 %/subject alignment coverage: 98.9 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF83184.1"
/db_xref="GI:9105206"
/translation="MTYNSILSIRINQACIQARKFLPFLHLPLVGVMLPFLGLMG
LAGDQWADLLRLSGGQWLLKQHWITEYLHVRGKWLIVAGLCLGLMLVRCRGLG
AGCYRLASLYVVVTLVSTLSILVKRLTGMDPCWDLVRYGGPPFGLPESRHLK
ASGCFPAGHASAGYAVWCLYFAAWGVCFAWRWGLMGLMIVAGLIFGISQOLRGHFLS
HDLWSLTICWLVAGLGFYLFVSPVTRVVSAQQRANT"
1479. 3137
/locus_tag="XF0375"
1479. 3137
/locus_tag="XF0375"
/note="similar to SP130845 (percent identity: 36 %/query alignment coverage: 98.0 %/subject alignment coverage: 98.9 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="inner membrane protein"
/protein_id="AAF83185.1"
/db_xref="GI:9105207"
/translation="MSFSTLVILVRRMSEFNRVRPVESTSVLGTSLFALACNT
MFWRSAMTSAGSISIGVLSALLVTYHALLGLVWVRWAKPLTLFLVITAFATHY
MNSYSYLDADLRNVFNDHRESLELTLSALIPLLFYAGPIAVLWLRFRQRPWS
RALGRMLLVIVVVGAGCMLSFOLSKALNDRVRLATPINYIMALKMSLNSD
SVLKPSKPIGEDAVATPPIPSRPLIAVIVGETARAQNLGNGVROTPPOLAON
DVINPDMHSCGTNTEVSPCMFSPYGRNRYDESKIRHQSLHLVLERISTLWRDN
QSGCKGVCLQQLQDNDKADPTLCSSGRCDMEILLKDFVQVRSKLGDVRVVVHQL
GSHGPSYFORYFVAFQFSPPTCNLGSCEQVVAAYDNSLLTYDHLFVLTIGLRL
DMSDYDTAMITVLSHGESIGSLGKGLYHGMFYAIPAEQTRVPMVWFQSKQFVQSRQID

gene

CDS

/note="similar to SP130845 (percent identity: 36 %/query alignment coverage: 98.0 %/subject alignment coverage: 98.9 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="inner membrane protein"
/protein_id="AAF83185.1"
/db_xref="GI:9105207"
/translation="MSFSTLVILVRRMSEFNRVRPVESTSVLGTSLFALACNT
MFWRSAMTSAGSISIGVLSALLVTYHALLGLVWVRWAKPLTLFLVITAFATHY
MNSYSYLDADLRNVFNDHRESLELTLSALIPLLFYAGPIAVLWLRFRQRPWS
RALGRMLLVIVVVGAGCMLSFOLSKALNDRVRLATPINYIMALKMSLNSD
SVLKPSKPIGEDAVATPPIPSRPLIAVIVGETARAQNLGNGVROTPPOLAON
DVINPDMHSCGTNTEVSPCMFSPYGRNRYDESKIRHQSLHLVLERISTLWRDN
QSGCKGVCLQQLQDNDKADPTLCSSGRCDMEILLKDFVQVRSKLGDVRVVVHQL
GSHGPSYFORYFVAFQFSPPTCNLGSCEQVVAAYDNSLLTYDHLFVLTIGLRL
DMSDYDTAMITVLSHGESIGSLGKGLYHGMFYAIPAEQTRVPMVWFQSKQFVQSRQID

gene

CDS

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF83186.1"
/db_xref="GI:9105208"
/translation="MRLSRITDFSYLLEKLRIRLDLQRIQAHYGGTLLDLVLITPE
LAEAYENL"
complement (3751. 4959)
/locus_tag="XF0377"
complement (3751. 4959)
/locus_tag="XF0377"
/note="similar to SP153554 (percent identity: 37 %/query alignment coverage: 94.3 %/subject alignment coverage: 95.9 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder/Start codon shift: -108"
/codon_start=1
/transl_table=11
/product="cytochrome P450-like enzyme"
/protein_id="AAF83187.1"
/db_xref="GI:9105209"
/translation="MKLTDLGNPAFLNPYLYETLRAQAFVSGPNALMTGRYSYL
DSLHNRMGKKYMEVRVYQSDAADMPLFAQSRMFTINPPATHLRLGVMOAFT
GRESMRPLAIDTAHQILDNFEQKPSVDLVAEPFPMQIICMOMVDIGDVTIG
IAVSKIAKVPDPSMSADELHASTAYBELAQYFKLIELRHLTGTDLSIFLRAE
DDEKLTDEIVSNVIMLIAGYETTSNMGNALIHRRHQLAKDLSLMPQAVS
ECLRYDGSVQFTMRAAMDDIEVEGELVPRGTVMFLMGAANDPQAFTPDQDITRK
QRLQSGFAGITHCLGYRLALIELECAUULLFERLPHRLAHLDALNWNQSRNLGRVN
TLIVDLHAKN"
complement (5841. 6545)
/locus_tag="XF0378"
complement (5841. 6545)
/locus_tag="XF0378"
/note="similar to GI5932372 (percent identity: 41 %/query alignment coverage: 80.8 %/subject alignment coverage: 79.4 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="thiamin-phosphate pyrophosphorylase"
/protein_id="AAF83188.1"
/db_xref="GI:9105210"
/translation="MEVSRIGSFQBIHDKTGTIVTPHPLLRPLMSQPRGIVLTIPDE
TDTARLIARTAPLLNGIYVLRNKLANTALTEQAQALLACROTGPLLLINDLEL
AQTIGADGVHLGHMDSNPISIAQAQIHPHAIIGVSCYQTERAKQAKAGASVGFQTF
YPSHTKTPPYRATPELLRQTTHLGVPRVAIGLTPKNTAPIEAGALLAVISGIYSA
KNPTALAKYQSQFNI"
6547. 6720
/locus_tag="XF0379"
6547. 6720
/locus_tag="XF0379"
/note="similar to SP1242453 (percent identity: 54 %/query alignment coverage: 93.0 %/subject alignment coverage: 98.1 %); identified by sequence similarity; putative; ORF located manually"
/codon_start=1
/transl_table=11
/product="rubredoxin"
/protein_id="AAF83189.1"
/db_xref="GI:9105211"
/translation="MMWFTWCTCGFLYRBEEGPPEPGIAAGTWNEDVSEBWCPCD
GIGKADFFMWTVG"
complement (6854. 7021)
/locus_tag="XF0380"
complement (6854. 7021)
/locus_tag="XF0380"
/note="hypothetical protein; identified by sequence

gene

CDS

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF83189.1"
/db_xref="GI:9105211"
/translation="MMWFTWCTCGFLYRBEEGPPEPGIAAGTWNEDVSEBWCPCD
GIGKADFFMWTVG"
complement (6854. 7021)
/locus_tag="XF0380"
complement (6854. 7021)
/locus_tag="XF0380"
/note="hypothetical protein; identified by sequence

gene

CDS

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="thiamin-phosphate pyrophosphorylase"
/protein_id="AAF83188.1"
/db_xref="GI:9105210"
/translation="MEVSRIGSFQBIHDKTGTIVTPHPLLRPLMSQPRGIVLTIPDE
TDTARLIARTAPLLNGIYVLRNKLANTALTEQAQALLACROTGPLLLINDLEL
AQTIGADGVHLGHMDSNPISIAQAQIHPHAIIGVSCYQTERAKQAKAGASVGFQTF
YPSHTKTPPYRATPELLRQTTHLGVPRVAIGLTPKNTAPIEAGALLAVISGIYSA
KNPTALAKYQSQFNI"
6547. 6720
/locus_tag="XF0379"
6547. 6720
/locus_tag="XF0379"
/note="similar to SP1242453 (percent identity: 54 %/query alignment coverage: 93.0 %/subject alignment coverage: 98.1 %); identified by sequence similarity; putative; ORF located manually"
/codon_start=1
/transl_table=11
/product="rubredoxin"
/protein_id="AAF83189.1"
/db_xref="GI:9105211"
/translation="MMWFTWCTCGFLYRBEEGPPEPGIAAGTWNEDVSEBWCPCD
GIGKADFFMWTVG"
complement (6854. 7021)
/locus_tag="XF0380"
complement (6854. 7021)
/locus_tag="XF0380"
/note="hypothetical protein; identified by sequence

gene

CDS

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="rubredoxin"
/protein_id="AAF83189.1"
/db_xref="GI:9105211"
/translation="MMWFTWCTCGFLYRBEEGPPEPGIAAGTWNEDVSEBWCPCD
GIGKADFFMWTVG"
complement (6854. 7021)
/locus_tag="XF0380"
complement (6854. 7021)
/locus_tag="XF0380"
/note="hypothetical protein; identified by sequence

gene

CDS

/note="hypothetical protein; identified by sequence

	similarity; putative; ORF located using Glimmer/RBSfinder"
	/codon_start=1
	/transl_table=11
	/product="hypothetical protein"
	/protein_id="AAF83190.1"
	/db_xref="GI:9105212"
	/translation="MLLKLKHNVDIFDKQSKLMTQKPSVHPKNKKSQKEIKLSHQKRD
	NNPRYKIFTYI"
gene	7207.. 9792
	/locus_tag="XF0381"
CDS	7207.. 9792
	/locus_tag="XF0381"
	/note="similar to sp P03815 (percent identity: 69 %/query
	alignment coverage: 99.5 %/subject alignment coverage:
	100.0 %; identified by sequence similarity; putative; ORF
	located using Glimmer/RBSfinder"
	/codon_start=1
	/transl_table=11
	/product="ATP-dependent Clp protease subunit"
	/protein_id="AAF83191.1"
	/db_xref="GI:9105213"
	/translation="MRMDKLSRFQNALADASLAVGRDHTIIEPVHFSALLDQGG
	STRPLMQAVNVLRLERLLEALPKVQTVNVSFNSLSLFRTKLAQOHG
	DFMASFVLAVDSDSGLGQALRAAGAEKKIEAIDKLRGGTFTVOTNAEORQA
	LEKYITDLTARAESGKLDVIGRDEIRTIQVLRRTKNNPVLIGEPGVGKTAIEG
Query Match	9.3%; Score 113.6; DB 1; Length 10092;
Best Local Similarity	47.5%; Pred. No. 28-23;
Matches 338; Conservative	0; Mismatches 374; Indels 0; Gaps 0;
QY	272 TTGTAGTGAAGACTTTATCGGTGAGCACTCGATCATCTGCTCTCAATGATTAAACAA 331
DB	4673 TGGTAATGAACGGTTCACCGGCGAGAGAGTGAATCAATGCGTCCCTTGGCATCGATA 4614
QY	332 ATGCGAAGAACTTGTAGCGCCTTATCTTGAAGAGGAAAGTAGTCATCGTCAATGATT 391
DB	4613 CGGCGCACCAATGATCGACAACTTTGAACAGAAACCATCCGTGATCTTGTGCTGAAT 4554
QY	392 TTGGAAGACGTTTGGGTTGGTGCATGATGACATGTCGCGGTGGTAAAGACGCC 451
DB	4553 TTGCGCTTTCCTTTCCCATGCAGATCATTTGCAAAATGATGATGGATATGGAGATG 4494
QY	452 ATGAAABAATCTGTAGTGCACAGCGGAGTTGCGGATTTTATCAGAGTACTCTCAAT 511
DB	4493 CGTCAACCTCGGTATAGCGGTGAGCAAAATTCGCAAGTATTCGACCTTCGCCAATG 4434
QY	512 CTCCTGAAGCGGGGCACATGTTGTTATGTCAGCGAAGACAGCTTTCCCAATACTTGATGC 571
DB	4433 CGGCCGATGAATTAGTACACGAAGCACTGCCTACGAAGAGCTTGGCAATACTTTACGA 4374
QY	572 CGGTCAATTAAGAACGTCGCGTCAATCCGGGATCAGATTTAATTTGATCCCTATGTACTT 631
DB	4373 AGTTGATTGAGCTACGCCGCCACCCACTCTCGGAAGTGAATTTCTATGTTCTCGCGTG 4314
QY	632 CTGATATCAAGCATGCGGCTGTGCGCAAGATATATCCGACCTGATCTTCTTAATGTGC 691
DB	4313 CTGAGAACAGCGCGAAGAACTAACCCATGATGAATCGTCTCCATGTGATATGTTAT 4254
QY	692 TGTTAGCCGCAACGGGCTGTGTAAGACGCTGGCACTGATGATCTACCAATTTGCTCA 751
DB	4253 TAATCGCGGTTACGAAACCAATCCAATATGATTGGCAATGCAATGATTGGTGGTGCATC 4194
QY	752 ACAATCTCAGCAGATGAATGATGTTTGGTGTGACCGGTTCTGGTTAGTTCGAGAGCCATTG 811
DB	4193 GCCATCCAGACCAACTCGCACTCCTCAAGAGTGAATCTGTCACTGATGCCACCAAGCGGAT 4134
QY	812 CGGAGACATTGGTTATAACCGCGGTTCACTGATTCGCGGCGAGCTGCCAAGATA 871
DB	4133 CGGAATGCTTACGTACGAGGTCGTTACAGTTTACAGTGCAGCGCGGCGATGATGATTA 4074
QY	872 CAGTGTCGCGCGGTATGGAATACAAAAAGATACGATGTTTTTTTGTATGATCGGTGCGG 931
DB	4073 TAGAGGTTGAAGCGGAGTTGGTTCCTCGTGGCACAGTAGTATTTTTTGTGCTGCTG 4014

QY	932 CTAACCGGACCTGAAGCATTGTAACAGCCTGAGCTGTTAATATTATTCATCG 983
DB	4013 CAACCGTGATCCGCGGCAATCTACTATCCGACCGACGCTGGATATTACTCG 3962
RESULT 10	
LOCUS	AE017008 300090 bp DNA linear BCT 16-MAY-2003
DEFINITION	Bacillus cereus ATCC 14579 section 11 of 18 of the complete genome.
ACCESSION	AE017008 AE016877
VERSION	AE017008.1 GI:29896738
KEYWORDS	.
SOURCE	Bacillus cereus ATCC 14579
ORGANISM	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus; Bacillus cereus group.
REFERENCE	1 (bases 1 to 300090)
AUTHORS	Ivanova,N., Sorokin,A., Anderson,I., Galleron,N., Candelon,B., Kapatal,V., Bhattacharyya,A., Reznik,G., Mikhailova,N., Lapidus,A., Chu,L., Mazur,M., Goltzman,E., Larsen,N., D'Souza,M., Walunas,T., Grechkin,Y., Pusch,G., Haselkorn,R., Fonstein,M., Ehrlich,D.S.D., Overbeek,R. and Kyrpides,N.
TITLE	Genome sequence of Bacillus cereus and comparative analysis with Bacillus anthracis
JOURNAL	Nature 423 (6935), 87-91 (2003)
MEDLINE	22608415
PUBMED	12721630
REFERENCE	2 (bases 1 to 300090)
AUTHORS	Candelon,B., Gailloux,K., Ehrlich,D.S. and Sorokin,A.
TITLE	The number of ribosomal RNA operons in Bacillus cereus
JOURNAL	Unpublished
AUTHORS	3 (bases 1 to 300090)
	Ivanova,N., Sorokin,A., Anderson,I., Galleron,N., Candelon,B., Kapatal,V., Bhattacharyya,A., Reznik,G., Mikhailova,N., Lapidus,A., Chu,L., Mazur,M., Goltzman,E., Larsen,N., D'Souza,M., Walunas,T., Grechkin,Y., Pusch,G., Haselkorn,R., Fonstein,M., Ehrlich,D.S.D., Overbeek,R. and Kyrpides,N.
TITLE	Direct Submission
JOURNAL	Submitted (12-MAR-2003) INRA, Genetique Microbienne, Domaine de Vilvert, Jouy en Josas 78352, France
FEATURES	1..300090
source	/organism="Bacillus cereus ATCC 14579"
	/mol_type="genomic DNA"
	/strain="ATCC 14579"
	/db_xref="ATCC:14579"
	/db_xref="taxon:226900"
	complement (109..660)
gene	/locus_tag="BC3070"
	complement (109..660)
CDS	/locus_tag="BC3070"
	/EC_number="3.4.21.89"
	/codon_start=1
	/transl_table=11
	/product="signal peptidase I"
	/protein_id="AAP10017.1"
	/db_xref="GI:29896739"
	/translation="MKENTKELFSWAKTIGFTLVLIATIRGLVFTPLVQGESWMP
	LKNRVLNKGYSISGLDRFDIIVPHKGYDLVKRVIPLPGDVEYKNDLVYNG
	KAMEPYLQKFKEKAVGRVLPDFTLEQITGKTVPQGVFVLGDNREVSKDGRMFG
	ISEEIVKGQAVFWPLEQVRAL"
	complement (684..1361)
gene	/locus_tag="BC3071"
	complement (684..1361)
CDS	/locus_tag="BC3071"
	/codon_start=1
	/transl_table=11
	/product="Copper homeostasis protein cutC"
	/protein_id="AAP10018.1"
	/db_xref="GI:29896740"
	/translation="MLEVIATCLEDVKRIESAGGKRIELISSYTEGGTTPSYAFIKKA
	VEVVISIPTHVMIRPHAKSFYVTEEBIEMKEDIIVAQKLAGAVVLGLVLRNERVED

gene	KLADLLSVVDGINVYTHRAIDDIKNPVEAMSTLKKFHKVTHVLVLTSGQGQNIVENIPVIV TEMQKVSQGQIQLVAGAGVTKENIKRLLDDEGTITQAHVGTAVREGKSCFSEIDPNLIVL ELBIIK"
CDS	complement (1560..2267) /locus_tag="BC3072" complement (1560..2267) /locus_tag="BC3072" /codon_start=1 /transl_table=11 /product="Transcriptional regulator, ArsR family"
gene	/db_xref="GI:29896741" /translation="MNVPNISIYAKLIAEPTRAILDCLMNQALPASELATMAKVS HPTISLSKLVEGNLLTVBOHGHRYRLANOEAVELEKLTGIAPTQVRSLKQSS QLKQIRYARTQYDLAGKLVGVIETKLLDRQFIILEDEGYIVTEQKKFLNFGINVD EADTKRVPAKPCLDMDSERRYVHISGWLGAIAKLLPQGWITKTDKNRAVHLTKKGVK LLEDOLGIDMKTEERKA"
CDS	2337..2807 /locus_tag="BC3073" 2337..2807 /locus_tag="BC3073" /codon_start=1 /transl_table=11 /product="hypothetical protein"
gene	/protein_id="AAP10020.1" /db_xref="GI:29896742" /translation="MNAIDLSILNKLKTRRSSEKLSNLPDNFLNWKPDPEAMSFQEM IRHVSSTFYVHMIKKNQSINDIRTPYDDDEPITCVKKEIELASHFTDPIGVVQSIS IAEDSLRLDSDVGYYQRYLGDMLLRITAYDAHVAGQFLQYLRMNLRLPLND"
misc_feature	complement (3023..4172) /locus_tag="BC3074" complement (3023..4172) /locus_tag="BC3074" /locus_tag="BC3074" /note="similar to Acyl-CoA dehydrogenase, short-chain specific; EC number 1.3.99.2"
gene	complement (4327..4959) /locus_tag="BC3075" complement (4327..4959) /locus_tag="BC3075" /codon_start=1 /transl_table=11 /product="EpiH/GdmH-related protein"
CDS	/protein_id="AAP10021.1" /db_xref="GI:29896743" /translation="WKENKKSKERIFQVELLMTCSAILVSVAAVDIKSYRFKAND VKTAGIVLGASNVNKPSPVFERINHAISLYXNGNIKLIITGGTKEAELEEAR ARVAMQGVKEEDILLETLSLFTENLKNAKQVHENGIOITITIVSDPLHMKRAMRI AKHNIAYASPSTPSTAYKTLDTIEIPFFFKELFSYIGYVTSPLKALKGD"
gene	complement (5289..5876) /locus_tag="BC3076" complement (5289..5876) /locus_tag="BC3076" /EC_number="2.3.1.-"
CDS	/codon_start=1 /transl_table=11 /product="Acetyltransferase"
gene	/protein_id="AAP10022.1" /db_xref="GI:29896744" /translation="MEIIRHAKURLMDNTDVTFLPSIVEGNRDIIWAYLISKMSDQV MOQYQKAIKGYRGAGTQPIPTVDDQSNKIVGSTRLYNISVEDKTVELGQTYHPSVQ RTSINTECKQYMLLOYAFKELHMLRVLQIKTDARNEKAQRAIERLCAVKEGVLRNERKLP NGYVRDAVYSIIISSEWPSVKLEKLELYKEKI"
CDS	complement (6268..7596) /locus_tag="BC3077" complement (6268..7596) /locus_tag="BC3077" /codon_start=1 /transl_table=11 /product="hypothetical Membrane Spanning Protein"
gene	/protein_id="AAP10023.1" /db_xref="GI:29896745" /translation="MSBIEUKSNVYIKTGREKGIILREILASLVGVFMFPVPTINGAS SIMIDHIVSTIRASVGPVYPVALFVNAIGALPYFTKKWNAISVDICFSILKVGWV SIMIDHIVSTIRASVGPVYPVALFVNAIGALPYFTKKWNAISVDICFSILKVGWV

gene	PGLIYCLKVGAMFFADIPGFLYEKLVISVSLNVPIGSAPFLALLVGYGLLEFIQTPT RPIRMIIITAKTLDINHLWNVFIVTLVVTFITAITVRYPPLSRKPDITYTEEGFPFV YKEWMLBRAYEDALLESAPSIMKNIAVNLLKDGFTMTGILGPSIMSVGLIGTVLAKF TPIDWLSIYFYPTETLIQOLPEADLAKAASVGIAGMELPSSLVWSAPLVTKFVIWV SVSSILSPESASICPLLSDTLPVKSELIIIVYQRTILTLLIITIPIAVLLL" complement(7708...8586) /locus_tag="BC3078" complement(7708...8586) /locus_tag="BC3078" /EC_number="2.3.1.-" /codon_start=1 /transl_table=1 /product="acetyltransferase" /protein_id="AAP10024.1" /db_xref="GI:29896746" /translation="MHPIINSLVEKLIQEQFPPEAHLEVKPVKPSGVGNRTFHLGDSE SVRLPSDVDAVEPAQKESNLVLINKGLSLPISTPLAKNPSEAPLPWPGINKWIEGE TVTQNVDVINEFAADLGSLFWELQSINASNGPLAGTHNFVRGLISVYDEARVAE NNKVDFEALKHNLVALNSTWDWRKPWWHGDAFCNILLVKDGLCAVIDFGILGVG BPCADAAMATFPDENRNFKVEVIRMDDETNNRARGNALWKALITYDANRESNKVA EESYRFQIVDDYER" complement(8587...9102) /locus_tag="BC3079" complement(8587...9102) /locus_tag="BC3079" /EC_number="2.3.1.-" /codon_start=1 /transl_table=1 /product="Acetyltransferase" /protein_id="AAP10025.1" /db_xref="GI:29896747" /translation="MERNISKEAIRIATENESDSIILLKEVAQWLQYKEVDQMOWL LGEATABIUECKREKYTVYLKEAEIVGITVSPKQNEDEHI FKGEEVSNSLYIHRI FAVRKYKGGINGEWILRWBEENVQCDEYKLLDCVGHNRITLANDFYKKNSFEYVYGSTD GLSKPFQKRMR" complement(9126...9590) /locus_tag="BC3080" complement(9126...9590) /locus_tag="BC3080" /codon_start=1 /transl_table=1 /product="hypothetical protein" /protein_id="AAP10026.1" /db_xref="GI:29896748" /translation="MKRGILVDILDGDVNLNLKVLKPIDITFDWRVRVDESYSILIARG ELDRLAFPPSVVGEVGLKKDKONTYYILIFADLKAYPKGEIADLETYEFPKFSKC EVWVLVDAQYIQIYAKDXDQSIEMLVENAVNQGFVEYITDENDGRTRLVSW"
gene	EWWVLVDAQYIQIYAKDXDQSIEMLVENAVNQGFVEYITDENDGRTRLVSW"
CDS	/locus_tag="BC3081" complement(9807...10244) /locus_tag="BC3081" /codon_start=1 /transl_table=1 /product="hypothetical protein" /protein_id="AAP10027.1" /db_xref="GI:29896749" /translation="MMTNSNYDYWRQNDKLISDIEKAINGEYSATSCYAKLANMAPN QVEQGQLHEIRNDEIKHFNVQIYLTNLRQRPQPQITEDCPNTYLOGLSFATQDEORC TVDFYLEISDETSVHLKELTRIAADEQNHWLFVLYFYFKSK" 10551..11057 /locus_tag="BC3082"
gene	/locus_tag="BC3082"

	Query Match	7.5%	Score 90.8	DB 1	Length 30090
	Best Local Similarity	64.0%	Pred. No. 6.5e-16		
	Matches 137	Conservative 0	Mismatches 77	Indels 0	Gaps 0
QY	916	TGATCATCGTGGCGCTAACCGGACCTTGAGACATTTGAACAGCGTGCAGCGTTTAAT	975		
Db	59764	TGATGATTGGTGCAACAATTGTGACTTTACGTTATTGAAAGCGCTGATAATTCAC	5970		
QY	976	ATTTCATCGGGAAGATCTTGATATCAAGAGCGCTTTTAGCGCGCCCGCCCGGCATCTCGCT	1015		

FOALHPWHAIIKHLNQPVVREAREMVELQSLRQLQIEGVQHRVSVRIKTPWSIY
NKMDNDNSFQWMDVDFGLVVKVSAVDCYHALGIVHGTLPKDFRFDFTAIPTWING
YQSLTFLFPGVIEQVIRTEEMDMIAERHAYKCSGSDSPNQAQRVHAWIA
ELLESORTAISOMELDKFVDLPFVDFVLPFKGI PALPKNSFALDPFAVAVHDIQ
NHAVSRVDKLMPLRDKLSGQVETIITARSALPKPOMLEFVSSKARTAIRHOLKO
LOHEDAVOLGHRMLDRALEMDSSLERLPKGRDLAPLAHEHFORLEALLADIALGNWM
PNQALULTYELLGNTPISHSKILINGSERGVISFANCCQPIGDVVMGYHTTG
KGIHVHRLDCPNLNRSLRPERWFDLNDLVNSGDYDASLLVVEAENRTGVLAQAAAI
AQSHSNIHVDYLERDFNAALLRFSIQVDRRHLAEVMMRLRLPAVQSVYRQ"
3592..3975
/locus_tag="XF0353"
/locus_tag="XF0353"
3592..3975
/locus_tag="XF0353"

gene

CDS

/note="similar to Sp|p40431 (percent identity: 54 %/query alignment coverage: 99.2 %/subject alignment coverage: 99.2 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder/Start codon shift: 180"
/codon_start=1
/transl_table=11
/product="translation initiation inhibitor"
/protein_id="AAF83163.1"
/db_xref="GI:9105181"

translation="MSFOTIHTKPAALGPYSQAVRVGNVYFSGQIPLDPATCTIV

VGDIAVQARRAFDNRARVAREANGSLKIVRLGLYLDLEQPAVVNVNMQEYFQAPFP
AKSTIQVGLPKGADFVDVAVMVID"
3983..6139
/locus_tag="XF0354"
3983..6139
/locus_tag="XF0354"

gene

CDS

/note="similar to Sp|p24230 (percent identity: 55 %/query alignment coverage: 97.8 %/subject alignment coverage: 101.3 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="ATP-dependent DNA helicase"
/protein_id="AAF83164.1"
/db_xref="GI:9105182"

translation="MPRARSHAPDLNVAGNAPLSVLPVPGVLAGKFAARGIFTLQDL

WLLPLRYEDRTRLPPIAELOGVAVQIEARVEAFERSFRYRPLRLVLTDDAHILT
MFFPHRAQIAFTVGTFRVAYGPKLGOGWELVHPSYILAPGEDVRLNDCIDPV
YPIEGVGPAIVQLIREALEHLPTDVALELLPDWLSDLGLPMLRTALLINHRPPLN
TDIARLMVGTQALSLLELLAHLQSLRRQIALQRHSPALPDGALVASLLHAL
PPLHAAQOQFAQAAADVAKVPMRLVQGGSGKTVAALAAALLAVEQNKQVALA
APTALLAEHNLRLDWLEPLGVVHLLAAKTGKARLQALNDVASGDAQVVGTHAL
MOESVHEDLATITVDEOHRFGVHORLLLRDKGATAGIVPHOLVMTATPIPTLMAA
YADLVSVIDENPGRTPQTVNLSAERPPELVERILACAGQVQVWCTILDESOT
EAPQPPSNDIEHRSVQAAQTTEALSAQLPVGVGLVHGRMKAEEKQRTNRAPKC
NEIDLVAFTVLEGVDPVNASIMTIENAEFRGLAQLHOLGRVGRSVSSCVLLYQ
PPLSLARQLRLEIMRQTDGFAIEKDLRGLGELLGTRQTGLAAFRVADILDAHL
LPRVYSLANMLLDESPQLADQVIARWIGSAVYASA"
6214..6360
/locus_tag="XF0355"
6214..6360
/locus_tag="XF0355"

gene

CDS

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF83165.1"
/db_xref="GI:9105183"
/translation="MSPGPTALNAGGVAVLLSDPTIAARLKSFALEKRVAAALFTGI
VLR"

gene

CDS

6723..7925
/locus_tag="XF0356"
6723..7925
/locus_tag="XF0356"

translation="MSPGPTALNAGGVAVLLSDPTIAARLKSFALEKRVAAALFTGI

6723..7925
/locus_tag="XF0356"
6723..7925
/locus_tag="XF0356"

Query Match

Best Local Similarity

Matches

Conservative

0; Gaps

0; Indels

0; Gaps

0; Indels

0; Gaps

0; Indels

0; Gaps

0; Indels

7113 GATCTGTGTACTGAATTTCTTTCCCTTCCATGCCGATCATTTGCGAGTATGAT 7172

436 CTGATAAAGAGACCATGAAAAATCTCTGAGTGGCACACGGAGTGTCCGCAATTTATC 495

7173 GTGGATATCAGCAGGCTATCAGTCTAAGTGTGGCGGTGAGCAACATTTGCCAAGTATTG 7232

496 AGGAGTATCTCTCAATCTCTTGAAGCGGGGACATTTGTTATGTGTGACGCAACAGCTT 555

7233 GACCTGCGCGGATGTCGCCGATGAATTTGGTGCACGCAAGTGTGCTTACGAAGAACTA 7292

556 TCCCAATCTTATGATCGCGGTCTAATAAGAACTGCTCGCTCAATCCGCGGATCAGATTAATT 615

7293 GGCATTACTTCAAGGTTGATCGAGTGCCTCCCGCGGACCTTGGAACTGACTTATT 7352

616 TCGATCTTATGATCTTGAATATGAAGGCATGCGGCTGTTCGGACAAGGATATATCTCGCA 675

7353 TCCATGTTGCTGCGTCTGAGGAGGAGGCGCAAGAACTGACGCATGATGAATCTGTGTC 7412

676 CTGATCTTATGATCTTGAATATGAAGGCATGCGGCTGTTCGGACAAGGATATATCTCGCA 735

7413 AATGTGATTTGTTTATGTTGCTGAGGAGGAGGCGCAAGAACTGACGCATGATGAATCTGTGTC 7472

736 ATCTACCATTTGCTCAACATCTCTGAGCAGATGATGATGTTTTCGGTACCGCTTCTGTTA 795

7473 TTGATGCTGTCATCGTCACTCCGAAGCAGCTTGCACGACTCAAGAGTATTTGCTGTG 7532

796 GTTCCGAGAGCATTGCGGAGACATTTGGTTTATAAACCAGCGCTTTCAGCTGATTCGCGG 855

7533 ATGCCGCAAGCATATTGGAATGCTTACGTTACGATGATCGTACAGTTCACGATCGC 7592

856 CAGCTGTCCCAAGATACAGTGTGCGGCTATGGAATCAAAAAGATACGATGTTT 915

7593 GCGCGCATGATGATGTCAGCATTTCAAGTGTAGTACCTCGTGGCAGCATTTGTTT 7652

916 TGTATGATCGTGTGCGCTTAAACGGGACCTGAAGCATTTGACACGCTGACGTTTAAAT 975

7653 TTGATGCTGTGCTGCTTAAACCGTATCGCGGCAATTCATCTGATCCGATCATCTGGAC 7712

976 ATTATCGGAGAGATCTTGGTATCAAG 1002

7713 ATCACCGGACGACGAGCTCCAG 7739

AL646080 190050 bp DNA linear BCT 11-APR-2003

Ralstonia solanacearum GM11000 megaplasmid, complete sequence;

segment 5/11.

AL646080 AL646053

AL646080.1 GI:17431113

Ralstonia solanacearum

Ralstonia solanacearum

Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;

Burkholderiaceae; Ralstonia;

1 Salanoubat, M., Genin, S., Attiguenave, F., Gouzy, J., Mangenot, S.,

Arlat, M., Billault, A., Brottier, P., Camus, J. C., Cattolico, L.,

Chandler, M., Choise, N., Claudel-Renard, C., Cunac, S., Demange, N.,

Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schiex, F.,

Siguer, P., Thebault, P., Whalen, M., Wincker, P., Levy, M.,

Weissenbach, J. and Boucher, C.A.

Genome sequence of the plant pathogen Ralstonia solanacearum

Nature 415 (6871), 497-502 (2002)

21681879

11823852

2 (bases 1 to 190050)

Boucher, C.A.

Direct Submission

Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston

Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie

Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,

BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean Dausset-CRPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS 118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA URGV, 2 rue Gaston Crémieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27, F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire INRA, BP27, F31326 Castanet-Tolosan Cedex
Christian.Boucher@toulouse.inra.fr
<http://sequence.toulouse.inra.fr/R.solanacearum.html>.

COMMENT

FEATURES

source

1..190050

Location/Qualifiers

/organism="Ralstonia solanacearum"

/mol_type="genomic DNA"

/strain="GMI1000"

/db_xref="taxon:305"

/plasmid="megaplaemid"

51..362

/gene="RSp0643"

/note="synonym: RS05579"

51..362

/gene="RSp0643"

/function="miscellaneous; unknown"

/note="Product confidence : hypothetical

Gene name confidence : hypothetical

predicted by Codon usage

predicted by Framed"

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="HYPOTHETICAL PROTEIN"

/protein_id="CAD17794.1"

/db_xref="GI:17431114"

/db_xref="GOA:Q8XS38"

/db_xref="TrEMBL:Q8XS38"

/translation="MKTSPSPNPAAGPASDPEAAGDQGVAFEQLHEYNHTITIRPR

RIGGLWRAFAVSLDRALVAVGTVGQAFRTPCAARDAVDAARSFIDARLKACRP

A"

397..555

/gene="RSp0644"

/note="synonym: RS05578"

397..555

/gene="RSp0644"

/function="miscellaneous; unknown"

/note="Product confidence : putative

Gene name confidence : hypothetical

predicted by Codon usage

predicted by Framed"

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="PUTATIVE SIGNAL PEPTIDE PROTEIN"

/protein_id="CAD17795.1"

/db_xref="GI:17431115"

/db_xref="GOA:Q8XS37"

/db_xref="TrEMBL:Q8XS37"

/translation="MERQLFAIVIAVSAVSAVSLARDLNDLDRDFPPGQALVHGH

ATASAGR"

complement(575..1135)

/gene="RSp0645"

/note="synonym: RS05577"

complement(575..1135)

/gene="RSp0645"

/function="miscellaneous; not classified regulator"

/note="Product confidence : probable

Gene name confidence : hypothetical

predicted by Codon usage

predicted by Homology

predicted by Framed"

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="PROBABLE TRANSCRIPTION REGULATOR PROTEIN"

/protein_id="CAD17796.1"

/db_xref="GI:17431116"
/db_xref="GOA:Q8XS36"
/db_xref="TrEMBL:Q8XS36"
/translation="MKRISTFPDFVDMKKVAIKFRSPAAAPPGQAALSALDRTDRAIL
KWLQRDASISNALADKVSUPACGLRRRVERLKGCIIRGVVVALLEPKALDGLTLVMI
GVILDRSTPESFADFERAKVQKVPACLECHVVVTGEFDYFMLVTRTRDNDNSNRLHABQLL
YLPVGRGIRITMTLKLQVLTSTQLPIA"

1244..2260

/gene="RSp0646"

/note="synonym: RS05576"

1244..2260

/gene="RSp0646"

/EC_number="4.1.99.4"

/function="miscellaneous; hypothetical/global homology"

/note="Product confidence : probable

Gene name confidence : hypothetical

predicted by Codon usage

predicted by Homology

predicted by Framed"

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="PROBABLE 1-AMINOCYCLOPROPANE-1-CARBOXYLATE

DEAMINASE (ACC DEAMINASE) PROTEIN"

/protein_id="CAD17797.1"

/db_xref="GI:17431117"

/db_xref="GOA:Q8XS35"

/db_xref="TrEMBL:Q8XS35"

/translation="MNLNKHPRHPLTTFGPTFIQPKRLSAHLGGKVELYAKRDCNSG

LAFGNKTRKLEYLVPELAGGYDTLVSIGIQSNQTRQVAARAHGLKCVLQENW

VNYADVYDRGNIELSLGADVLDAAGFDIGIRPSWEQAMEDVRAGGKPPFIPA

GCSBPLGGLGVGFABEVROQAEFGFRFDYIVVCSVTGTOAGMVFVFAADGRADR

VIGDASAKPQTRQILRIARDTAKVELGRDITDDVDLDTYXGPEYGLPNEGTL

EAILRCARQEGMLTDPVTEGKSMGMDIRVRGGEFFPEGSRVLYIHLGGVPALNAYSFL

FRNG"

complement(2336..3508)

/gene="RSp0647"

/note="synonym: RS05575"

complement(2336..3508)

/gene="RSp0647"

/EC_number="4.2.1.17"

/function="miscellaneous; hypothetical/global homology"

/note="Product confidence : putative

Gene name confidence : hypothetical

predicted by Codon usage

predicted by Homology

predicted by Framed"

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="PUTATIVE ENOYL-COENZYME A HYDRATASE PROTEIN"

/protein_id="CAD17798.1"

/db_xref="GI:17431118"

/db_xref="GOA:Q8XS34"

/db_xref="TrEMBL:Q8XS34"

/translation="MSALPKPAEPAAAPAEVLFHVNRVAILTLNRPALNALSHGM

VRQSALLAQCRDDPRIAAVLRGNGKFCAGGVRALYRLARQGHQPDGWLQFF

VDEYRLDYALHRFPKPVVIGLDTGAMGGMLGQGRARLVVTERSRAMPETRIGFLP

DVGATHFLGAMPPEMALYIGLTGTLGADALHCLGADVCPAEWLASFEDLLRLSH

AGDPGLMQALRVFPPCNTVPHAAIARTTPIVIRYDFRSTIDRIVATLESLEDR

PAREARWLQATLDAMATHSTMLHVTREALLRGQLTAEFCRMELGIVARAIIEGD

FCEGVRAHLVDKDRPGWAPATLAGVRERVVRHFLSSPWGAAHFLAHLGAQPQAA"

complement(3505..4305)

/gene="RSp0648"

/note="synonym: RS05574"

complement(3505..4305)

/gene="RSp0648"

/EC_number="4.2.1.17"

/function="small molecule metabolism; degradation; carbon

compounds"

/note="Product confidence : probable

Gene name confidence : hypothetical

predicted by Codon usage

[illegible]


```
OY 496 ACAGATATCTCTCAATCTCTGAAGCGCGGGCACTTCGTTATGGTGCGAGCGAACAGCTT 555
Db 180936 GACCCCTGCGCGGATGTGCGCGGATGAATGGTGCAGCGCAAGTCTGCCCTACGAGAACTA 180877
OY 556 TCCCAATCTTGTATGCGCGGTCTTAAGAAGCGTGCCTCAATCCGGGATCAGATTTAAAT 615
Db 180876 GCGCATTAATCTCAAGGTTGATGAGTTCGCGCGCGCGGATGCACTGCACTGATTTGATT 180817
OY 616 TCGATCCTATGTAATCTCAATATGAAGCATGCGCTGTCGACAAAGGATATACCTCGCA 675
Db 180816 TCCATGTTGTCGCTGCTGAGGAGGAGGCGCAGAACTGACGCATGATGAATCGTGTC 180757
OY 676 CTGATTTCTTAATGCTGTTAGCGGCAACCGAACCGGCTGATAGAGCGTGGCACTGATG 735
Db 180756 AATGTAATTTGTTGTTATTTGGGTGCTACGAGACCACTCCAACTGATGGCAATGCA 180697
OY 736 ATCTACCAATTTGCTCAACATCTGAGCAGATGAATGATGTTTGGCTGACCGTTCGTTA 795
Db 180696 TTGATTCGGTTGATCGTCATCCGAGCAGCTTGCGAGCTCAAGAGTGAATTTGTCGCTG 180637
OY 796 GTTCCGAGAGCCATTCGGAGACATTTGCGTTATAAACCCCGGTTTCAGCTGATTCGCGG 855
Db 180636 ATGCCCGAAGCGGATTTGGAATGTTTACGTTACGACGGATCAGTACAGTTCAAGATACG 180577
OY 856 CAGCTGTCCCAAGATACAGTGTGCGCGGTATGGAATCAAAAAGATAGGATGTTTTT 915
Db 180576 GCGGCGATGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 180517
OY 916 TGTATCATCGGTGCGCTACCGGACCTGAGACATTTGAACAGCCTGACCGTGTGTTAT 975
Db 180516 TTGATCTGTTGCTCTCAACCGTATCCGCGCAATTCAGGATCCGACCAATTTGGAA 180457
OY 976 AT 977
Db 180456 AT 180455

RESULT 14
LOCUS AE017236/c 302325 bp DNA linear BCT 02-FEB-2004
DEFINITION Mycobacterium avium subsp. paratuberculosis str. k10, section 10 of
16 of the complete genome.
ACCESSION AE017236 AE016958
VERSION AE017236.1 GI:41396889
KEYWORDS
SOURCE
ORGANISM
Mycobacterium avium subsp. paratuberculosis str. k10
Mycobacterium avium subsp. paratuberculosis str. k10
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
avium complex (MAC)
REFERENCE 1 (bases 1 to 302325)
Li, L., Bannantine, J., Zhang, Q., Amonsin, A., Alt, D. and Kapur, V.
Direct Submission
Submitted (05-SEP-2003) Biomedical Genomics Center, University of
Minnesota, 1971 Commonwealth Ave., St. Paul, MN 55108, USA
LOCATION/Qualifiers
1. .302325
/organism="Mycobacterium avium subsp. paratuberculosis
str. k10"
/mol_type="Genomic DNA"
/strain="k10"
/sub_species="paratuberculosis"
/db_xref="taxon:262316"
156..2294
/locus_tag="MAP2433"
156..2294
/locus_tag="MAP2433"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA04750.1"
/db_xref="GI:41396889"
/translation="MRTARRRRTRADRNVLPGRVEIDNVQPVVSCGTYPAKAVVG
```

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

```
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAS04754.1"
/db_xref="GI:41396893"
/translation="MTTDOVDARQVLAATVAIDHVGIAVADLDAALAWYHDHIGMI
LVHSEVNEQGRREAMLAVRGAPGTAQQLMAPIDDSSTIAKFLDKRGPGIQQWAMR
LSDTLTLERLEQGVRLIYDAPRRGTANSRINFHPKADGGVLIELVEPA"
complement(7390..7686)
/locus_tag="MAP2438c"
complement(7390..7686)
/locus_tag="MAP2438c"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAS04755.1"
/db_xref="GI:41396894"
/translation="WARRTPRRQRLLGPPPALRRVEVPGDGHYEVRRPIAARAVK
TVRCPCDHEIRSGAHLVWVPEPTLGGAEDRRHWTGCAHRAHTRGPTRRWS"
complement(7710..8390)
/locus_tag="MAP2439c"
complement(7710..8390)
/locus_tag="MAP2439c"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAS04756.1"
/db_xref="GI:41396895"
/translation="MSRVLRLVIAQCTVDYVGRLLTAHLPASARRLLLFKADGSVSHADD
RAYKPLNMSPPCWLREAGDAAPVWVENKAGEOLRIIVEDIHDSHDLGVDPGLV
KDGVAHLQALLAEHVQLLAGEVTLVRREYMTAIGPVDLLCRDRGGSVAVIKRGE
IDGVQLFRYLELLNRDSVLAVRGVFAAQIIPKQARTLATDRGIRCVTLIDDKRGM
DSEYKRLF"
8427..10061
/locus_tag="MAP2440"
8427..10061
/locus_tag="MAP2440"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAS04757.1"
/db_xref="GI:41396896"
/translation="MSATOSTAQRIQGRVLEKIROSGRLPPTPAYGSLLLGRVTSQH
RRRIQIIMTVMLGANLIGGAALLIVIVAIPSPVENDAPANITFEGVPAYVASA
LAVGIWTRITVLRWAIEBKPTPDERTNIFLAPRVALVDLVMGAGLLTLVL
RTWDTMFIPRFVFGVCGVLATACYLALFAIRPVAQALEARGPPPRRUTAGIMG
RMTVMFLSGVPVFIAGLLALFEIWLRLTETEFAGVLIIVSTAALIIPGCLIMILA
WLTATPVRVRAALKRVERGDLRVDPGTGELGELQRFNAMDGLRERERVRDLF
GRHVREVALAARERPKLGRERHVAVFIDIVGSTOLVTRPPPAEVVSVLNQFPI
VVEVDHRCGLNKEGDATLAIFGAPNHLDCPEDAALTAARAIADRLANENPQCRAG
IGVAGQVAGNVGAREFEYTVIGEPVNEAARLCELAKSHSRLLATGDATGASEK
ERARSLGPTVLRGHERPTRLASPAGAADRPAQTD"
complement(10525..11787)
/locus_tag="MAP2441c"
complement(10525..11787)
/locus_tag="MAP2441c"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAS04758.1"
/db_xref="GI:41396897"
/translation="MPARSMVMPHCPVAHQAGTRCYRMTAALCGAGLASFAAMYCS
QALLPALSHYRIGPATAALVSLITGALALSIIPASVUSERYGRIRVMILISGVASSV
IGLLPFLPSGLVLLFGRAQGVAGLAGIPAVAMALLAEEDASSIGSAMGRYIAGTTI
GGLAGRIVPSVQVQTVALLACSLITLAGTAVFAVLVPRSRFTPKPASPVAALR
NLAGHLRNPVLAKLFAVGFLVMGGFVTVYNLYRAARPFGLASVVGLLFLLVLG
TGTSVVAGLRDRGRPVLGAALPIAVAGLLITVPATIAAIVAGVFTGGFFAHT
VASGWGAQDRDRAASALYLSFYLLGSVAGAGVGLYGVGWSATVCFVVVLLMA
GAALVALLVRDNGFRIGRVRVSVASVK"
11854..12738
/locus_tag="MAP2442"
11854..12738
/locus_tag="MAP2442"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAS04759.1"
/db_xref="GI:41396898"
/translation="MAVGDYEMFITTAEQLHVTAAEQLHVAQPTLTRMLARLEORLG
VALPDRHRRLLSLNTGYRIFYEHARRAQLELDSARRAIALADLAPVAGVLYLSFG
STVPLRLTARFKESSPRVTFTLEKGAESIADRVRSGGVDIGVSPRPVVERTLAWRSL
FRQLGVAVPDGHRFARGAANVMDLADPEFTVHPHFGMRLELDELCAAAQFRPRVA
LQAPNLTAASLVAAGLIGLSLVPIDGSSYFSGSVSKPLADADAYRDVGMWDSGRPLP
RSARDFIAAALAVRGGM"
12825..14348
/locus_tag="MAP2443"
12825..14348
/locus_tag="MAP2443"
/codon_start=1
/transl_table=11
/product="Alka"
Query Match 7.0%; Score 85.2; DB 1; Length 302325;
Best Local Similarity 49.8%; Pred. No. 3.7e-14;
Matches 216; Conservative 0; Mismatches 218; Indels 0; Gaps 0;
QY 552 GCTTTCCCAATACCTGATCGCGTCATTAAAGAACGTCGCGTCAATCCGGGATCAGATTT 611
Db 252449 GCTGTGCGACTACTTCGGCGGATCCTGGCGCGCGCGCGAGAGCCCAAGGACGATCT 252390
QY 612 AATTTTCGATCTTATGATCTTCTGAATATGAAGGCATGGCGTGTGCGGACAAAGGATATACT 671
Db 252389 GATCAGCGGCTGGCCAGCGGAGATCGACGCCGAGAAGCTGGAAGACGAGGAGATCTA 252330
QY 672 CGCATCGATTTCTTAATGTCTTTAGCCGCAACGAAACCGGCTGATAAGACGCTGGCCT 731
Db 252329 CTCCTTCTCGCGCTGCTGCTCGCGCGCGGGTGAGACGACGTACCGGCGCTGGGCGAG 252270
QY 732 GATGATCTACCATTTGCTCAACAATCCTGAGCAGATGAATGATGTTTTCGCTGACCGTTC 791
Db 252269 CTTGCTGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 252210
QY 792 GTTAGTTCCGAGAGCATTGCGGAGACATTTGCGTTTATAAACCGCGCGGTTTTCAGCTGATTC 851
Db 252209 GCTGCTGCGCAGGCCATCGAGAGGGGTGCGCTGGGAGCCCGCGCTGCTGACCATCAC 252150
QY 852 GCGGAGCTGTCCCAGATACAGTGTGTGCGCGGTATGGAATACAAAAGATACGATTTGT 911
Db 252149 CCGGCTGGCCACCCCGGACACCGAACTCGCGGGGTGCGCGATCCCGCGCGGCTCCACGGT 252090
QY 912 TTTTGTATGATCGTGGCGCTAACCGGACCCCTGAAGCATTTGAACAGCCTGACGCTGT 971
Db 252089 GATGCGGATGCTGGGCGCGCCCAACCGGAGAGGACCGCTACCCCGACCCGACAGGTT 252030
QY 972 TAATATTCATCGG 985
Db 252029 CGACATCTCCGCG 252016
```

```
RESULT 15
AF015825
LOCUS AF015825 35739 bp DNA linear BCT 07-MAY-1998
DEFINITION Bacillus subtilis 168 cot-rapA region sequence.
ACCESSION AF015825
VERSION AF015825.1 GI:2612880
KEYWORDS
SOURCE
ORGANISM Bacillus subtilis
Bacillus subtilis
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE 1 (bases 1 to 35739)
AUTHORS Rivoita,C., Soldo,B., Lazarevic,V., Joris,B., Mauel,C. and
Karamata,D.
TITLE A 35.7 kb DNA fragment from the Bacillus subtilis chromosome
containing a putative 12.3 kb operon involved in hexuronate
```

catabolism and a perfectly symmetrical hypothetical

catabolite-responsive element

Microbiology 144 (Pt 4), 877-884 (1998)

98240225

PUBMED

REFERENCE

AUTHORS

2 (bases 1 to 35739)

Rivolta,C., Soldo,B.,

Karamata,D.

Direct Submission

Submitted (24-JUL-1997)

Microbiennes, Rue Cesar-Roux 19,

Location/Qualifiers

1. .35739

/organism="Bacillus subtilis"

/mol_type="genomic DNA"

/strain="168"

/db_xref="taxon:1423"

/map="cotT-rapA region"

complement (<1. .129)

/gene="cotT"

complement (<1. .114)

/gene="cotT"

/codon_start=1

/transl_table=11

/product="spore coat protein"

/protein_id="AAC46305.1"

/db_xref="GI:2612881"

/translation="WDYPLNEQSFQITPYDRQPVYVPRPRPFVPPVYYP"

complement (122. .129)

/gene="cotT"

355. .1806

/gene="yjeA"

355. .361

/gene="yjeA"

368. .1771

/gene="yjeA"

/note="similar to Rhizobium leguminosarum NodB protein"

/codon_start=1

/transl_table=11

/product="NodB-like protein"

/protein_id="AAC46306.1"

/db_xref="GI:2612882"

/translation="MLAKRIKFWHLVIAVCVGLIGLPHNHLKKETVMNKVRTDSQ

YGNVRIATLVNDGKTFNVAIVYVFNKNEKMSALKRFAEVRQFQKTDVQDHTT

KENELNVYKIYVYAKQTVAVIYFVNGYKIGGAHQTVKTKTFNSKQAFISIDIFK

EDADYLKLSLAYHELKKNKDIAADALLKEGTAPKENFSRFAIKEDYIELYFDY

QVAGVLCQSTAIKKSLKDLKQYIDKAKNKKIKQKPKHEVLSLPKEETVDPN

QVIALTFDDGNPATNQILDSLLKKYKGHATFFVLGSRVQYYPETLIRMLKEGNEVG

NHWSHPLLRUSVKEALKQINDTDIIEKISGYRPTLVPRPPYGINDELRSQMDYV

ALMDVDPEDWDRNKTKIYDVRVMOAGDGRITLIHDIVRTSADAADETIKKLTQDQYQ

LVTVSQLEVKQREAK"

1775. .1806

/gene="yjeA"

complement (1811. .2298)

/gene="yjeA"

complement (1811. .2284)

/gene="yjeA"

/codon_start=1

/transl_table=11

/product="unknown"

/protein_id="AAC46307.1"

/db_xref="GI:2612883"

/translation="MKRLPMKASLVLFVAVFVAVKGAFAKAEHAYDGKSPYNDCA

SSGSTKSSNLVNASNOVIGVVELKFSSTCKTAWAKITWNNTLTSGYANARITNTD

GKRYNCDSAGGNGKAVAGKSCYTPMVDLDRPTSYAFGKYSGPNLNVWATTGSI"

complement (2291. .2298)

/gene="yjeA"

complement (2363. .2589)

/gene="yjeB"

complement (2363. .2399)

/gene="yjeB"

complement (2409. .2576)

/gene="yjeB"

/codon_start=1

/transl_table=11

/product="unknown"

/protein_id="AAC46308.1"

/db_xref="GI:2612884"

/translation="WDIPALSVAMHQASLAQNVIATLTKRLDTAQONADOTLKMIQH

FTLGQIIVKKA"

complement (2581. .2589)

/gene="yjeB"

2689. .3632

/gene="yjeC"

2689. .2696

/gene="yjeC"

2703. .3632

/gene="yjeC"

/codon_start=1

/transl_table=11

/product="unknown"

/protein_id="AAC46309.1"

/db_xref="GI:2612885"

/translation="MNLVMEKTFEQEYKLFMSMEQKREDVFRYTMRPPEKMTAIQV

PLKAKEPNGYDVMAAKMLGYLDVDAESGQKALQILKESHVSETAESALROCIPFAE

REOLRVNAKEIKFGLYIADPNKLOIQKGYCGFGGIPGTHWVNPNSYALPRIPSTIA

HEFHNVRFSYIDFHGVSVCYDYLVIISGLAESPAFELGTERLGPWYTRFDHEDLOY

SIDVINEVDKGFSEVSKYMGFDPIANDQGFAPAGLSAFAGIAGYTHAVQSFNNQHH

ITISEATRLDAKTIISQCGLFLHEKSVITYSGAF"

3602. .3630

/gene="yjeC"

complement (3610. .4024)

/gene="yjeA"

complement (3610. .4008)

/gene="yjeA"

/codon_start=1

/transl_table=11

/product="unknown"

/protein_id="AAC46310.1"

/db_xref="GI:2612886"

/translation="MKNRWIYAVETILIIIGLGSRAFSSVLPDTLNTYLGDSLWAA

MIFTGCGFLFRKLMTITGIISLFCFVIEFSQLYHAEWIDQIRDTSLGSLVLGYGFL

WSDIEAYTIGIAACAAIIBLLVLGKIKKRCM"

complement (4020. .4024)

/gene="yjeA"

complement (4084. .4697)

/gene="yjeB"

complement (4084. .4099)

/gene="yjeB"

complement (4109. .4684)

/gene="yjeB"

/codon_start=1

/transl_table=11

/product="putative lipoprotein"

/protein_id="AAC46311.1"

/db_xref="GI:2612887"

/translation="MKTWMSAITAAAVTSCFTGGAASFAPAKAAQAQTNLSENTN

QSAABLKVNLYNTAYKMPQQAQGLTINKSTKGDVHAAPGEPFVPGDNRFDLYHW

NMGQGFYSYHKDNTIIEIRYFGTGVRLNLGGVTEVLQKLGPNVRVLTVPFTD

BIDYVYDGRYELHFEVIGTDQTADHVNLLKAK"

complement (4690. .4697)

/gene="yjeB"

4814. .7787

/gene="yjeC"

4814. .4822

/gene="yjeC"

4830. .7787

/gene="yjeC"

/note="similar to formate dehydrogenase alpha subunit of

Moorella thermoacetica; member of the 4Fe-4S ferredoxins

family"

/codon_start=1

/transl_table=11

/product="formate dehydrogenase alpha subunit homolog"

/protein_id="AAC46312.1"

/db_xref="GI:2612888"

/translation="MAGKKTITINGVMEABEEQTVLQLNNSIEVQVCYHPSLGP
IETCDCTIVSINGELKRSCELBKGDVIDITLSPDKKQAVIYDQKILYHLYCTVC
DYNNGCBIHTVPMKSHINHOIIPFDHKPYKHDSHPFVYDPDQICILGRCVBAQD
VQVETIITIDIERKEPRVIMWNPVINESSCVSGHCSTVCECNAMKEKMEGEGYL
TOINNETIUMIEITKGVETGSGSLAIDSMESAMRDERIKKTKVTCTYCGVGSFDV
WTKGHDILUKVEPQEEBANGISTCVKFGWDFVNSEBELTKPLIREGDFHFEABWEE
ALLIASFTELKEAFGDSLAFITSSKCTNEESYLMOKLARGVIGTNNVDCNSRYCQ
SPATAGLFRITGVGDSGSIITDIAQDLVLIIGSNTSESHPVLSRIKRAHKLRQKV
IVADIRKHEMARERDLFQPRAGSDIVMLNATAKVLIENTKADERFLRERVNGRDEVV
KSLAPYITLEYAEKGTIDQETLIQMAEMIQDASVCAWAMGVQTHIGGSDTSTAIN
LLLVNTGNYGPKAGSPLRGNHNVGASDFGSMRDLRPFYEKVTDEQVRQKTERVMGV
PLPKEPGMTHNMBIKIHSQKQKAMYVRKEEMGLVDSNINHVAAYEKLDFFVQDIF
LSRTAEFADVLPASPSLEKEGTFTNTRERIQRLYQVPEPLGESKPDWQIIMEVANKL
GAGWLEHPADIMEBAAKLSPIYAGVTVYERLEGVNSLOWPNADKDSPLLFTERFPF
PDGKAILYPVOMTEPKERCEYDIIHVNNGRLLEHFEHGNLTVKSGISEKTPVEFLEI
SPELAERGIQDCTVRLITSPFGNVKVKCLITDRVKGEVILPMDNSEAAILNLTGS
HADKDTDPAYKTESAKMEILKHDGSLPLPKINHNGNPPQFQIGVQVHKWARKDIIF
PGDAVKRGMGNHG"
7762..8373
/gene="y1gD"
RBS 7762..7769
/gene="y1gD"
CDS 7780..8340
/gene="y1gD"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC46313.1"
/db_xref="GI:2612889"
/translation="MAKAIKRIQKIEVTEEDQQRDLREIEDALIDHKEAILETLHML
GHMNERGLVPLLRGLFGQDKVLDILVKKADTEANTLKNLLILFLGLMDLVQKLE
PLILKNAGVASAVEQKSEKTKGYFDIIRSLKDPINKSIITLLFSFLKGMQDITKEL
ERTQPPHQKHQEPREKRNKRD"
8347..8373
/gene="y1gD"
8525..9223
/gene="y1ha"
RBS 8525..8530
/gene="y1ha"
CDS 8537..9178
/gene="y1ha"

Query Match 6.5%; Score 78.8; DB 1; Length 35739;
Best Local Similarity 45.4%; Pred. No. 2.3e-12;
Matches 327; Conservative 0; Mismatches 387; Indels 6; Gaps 1;
QY 270 AATTGTAGTGAAGCTTTATCGGTGACGCACTGGATCATCTGTCTCCATTGATTAACA 329
Db 10864 AGTCGTGAACAAGCCCTTTACTCCGCGCGTGTATGAAGCAATGGAAACCGAGAATCAAGA 10923
QY 330 AAATGCGAAGAACTTGTAGCGCTTATCTTGAAGAGGGAAGTGTCTGTCAATGA 389
Db 10924 AATCACAGATGAACCTGATCAAAAATTTTCAGGGCGCGAGTGTGACCTTTGTTTCAAGA 10983
QY 390 TTTTGGAAAGAGCTTTGCGGTGTGCGTCAAGATGGATGCTCGGCTCGATAAAGAGA 449
Db 10984 TTTTTCATACCGCTTCCCGCTTATGTGATATCTAGCTGTGGAGTGCCTTCAGCGCA 11043
QY 450 CCATGAAAAAATCTCTAGTGGCACAGCGAGTTGCCGATTTTATCACGAGTATCTCTCA 509
Db 11044 TATGGAACAGTTTAAAGCATGGTCTGATCTTCTGGTCAGTACACCGAAGATAAAGTGA 11103
QY 510 ATCTCTGAA-----GCGGGGCAATTCGTTATGGTGAGGGAACAGCTTTCACAATA 563
Db 11104 AGAAGCTGAAAAAGCCCTTTTGGAGAACGAGATAGTGTGAGGAAGAACTGGCCCGCTT 11163
QY 564 CTGTATGCCGGTCATTAAAGACGTCCGCTCAATCCGGATCAGATTTAATTTCCGATCCT 623
Db 11164 TTTTGGCGGCATCATAGAGAAAGCGAAACAAACCGGAAACAGGATATTATTTCTATTTT 11223
QY 624 ATGTACTTCTGAATATGAAGGCATGGCGGTGTGGCAAGAGGATATATCTCGCACTGATCT 683
Db 11224 AGTGAAGCGGAAGAAACAGCGGAGAAGCTGTCCGGTGAAGAGCTGATTCGCTTTTGCAC 11283

Search completed: December 10, 2004, 16:12:22
Job time : 5416.29 secs

QY 684 TAATGTGCTGTTAGCCGCAACGGAACCGGCTGATAAGACGCTGGCACTGATGATCTACCA 743
Db 11284 GCTGCTGCTGCTGCGCGGAATGAACCACTACAAACCTGATTTCAAATGCGATGTACAG 11343
QY 744 TTTGCTCAACAATCCTGAGCAGATGAATGATGTTTTTGGCTGACCGTTGTTAGTTCCGAG 803
Db 11344 CATATTAGAAAACGCCAGCGCTTTACGAGGAACTGCGCAGCCATCCTGAACTGATGCCTCA 11403
QY 804 AGCCATTGCGGAGACATTCGCTTATAAACCGCGGTTTCAGCTGATTCGCGGSCAGCTGTC 863
Db 11404 GGCAGTGGAGGAAGCCTTTGCGTTTCAGAGCGCGCGCCCGGTTTGTAGGCGCAITGCCAA 11463
QY 864 CCAAGATACAGTGGTTCGCGCGGTATGGAATCAAAAAAGATACGATTTGTTTTGTATGAT 923
Db 11464 GCGGATACGGAGATCGGGGGGCACCTGATTAAGAAAGTGATATGTTTGGCGTTTGT 11523
QY 924 CGGTGCGCTAACCGGGGACCTTGAAGCATTTTGAACAGCCTGACGTTGTTTAATATTATCG 983
Db 11524 GGCATCGGCAAAATCGTGATGAAGCAAAAGTTTGACAGACCGCACATGTTTGTATATCGCG 11583

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2004, 11:14:31 ; Search time 672.942 Seconds
(without alignments)
9477.854 Million cell updates/sec

Title: US-10-627-124-1
Perfect score: 1215
Sequence: 1 atgagccaatcgattaaatt.....tcgtgcgtttgacggggca 1215

Scoring table:
IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_23Sep04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002as:*
- 7: Geneseqn2002bs:*
- 8: Geneseqn2003as:*
- 9: Geneseqn2003bs:*
- 10: Geneseqn2003cs:*
- 11: Geneseqn2003ds:*
- 12: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1215	100.0	1215	12	ADJ62038
2	617.2	50.8	1208	6	ABK74890
3	247.4	20.4	1224	10	ACF70115
4	247.4	20.4	110000	10	ACF65385.2
5	247.4	20.4	110000	10	ACF67367_31
6	82	6.7	1204	3	AA95667
7	81.4	6.7	1188	9	ACF06114
8	78.8	6.5	1191	3	AA95666
9	78.8	6.5	1221	3	AA95668
10	77.2	6.4	1221	3	AA95669
11	72.4	6.0	966	6	ABK74899
12	67.8	5.6	110000	4	AAI99682_39
13	67.8	5.6	110000	4	AAI99683_39
14	64.8	5.3	1213	6	ABK74891
15	61.8	5.1	110000	4	AAI99682_08
16	61.8	5.1	110000	4	AAI99683_08
17	59.2	4.9	25681	8	ADJ53165
18	55.6	4.6	1200	8	ACA38018
19	55.6	4.6	84428	12	ADM45913
20	55.2	4.5	1233	10	AAD55813
21	55.2	4.5	60196	10	AAD55810

22	55	4.5	1155	6	ABK74894	Abk74894 Bacillus
23	54.2	4.5	36538	10	ABV75558	Abv75558 Saccharop
24	53.6	4.4	1215	9	ACF06133	Acf06133 Bacterial
25	53.6	4.4	110000	4	AAI99682_08	Continuation (9 of
26	53.6	4.4	110000	4	AAI99683_08	Continuation (9 of
27	53.4	4.4	1400	2	AAQ14548	Aaq14548 Eryf gene
28	52.8	4.3	5880	4	AAI68687	Aai68687 Pseudomon
29	52.8	4.3	11355	4	AAI68692	Aai68692 Pseudomon
30	51.6	4.2	2000	8	ADA71938	Ada71938 Rice gene
31	50	4.1	17596	12	ADI14148	Adi14148 M. megalo
32	50	4.1	47981	4	AAF30757	Aaf30757 Micromono
33	49.6	4.1	1191	9	ACF06115	Acf06115 Bacterial
34	49.6	4.1	6885	2	AAI70153	Aai70153 S.longisp
35	48.2	4.0	9024	12	ADI14147	Adi14147 M. echino
36	47.6	3.9	1209	8	ABX56043	Abx56043 M. echino
37	47	3.9	1230	10	ADC36274	Adc36274 Weed cont
38	47	3.9	1449	10	ADC36279	Adc36279 Weed cont
39	46.6	3.8	1290	5	AAH65582	Aah65582 C glutami
40	46.6	3.8	1407	4	AAF71749	Aaf71749 Corynebac
41	46.6	3.8	1414	10	ADD13338	Add13338 C. glutam
42	46.6	3.8	2000	8	ADA71938	Ada71938 Rice gene
43	46.6	3.8	34980	5	AAH68525	Aah68525 C glutami
44	46.2	3.8	1248	9	ACF06136	Acf06136 Bacterial
45	44.8	3.7	1227	10	ADC36139	Adc36139 Weed cont

ALIGNMENTS

RESULT 1
ADJ62038
ID ADJ62038 standard; DNA; 1215 BP.
XX
AC ADJ62038;
XX
DT 06-MAY-2004 (first entry)
XX
DE Bacillus subtilis cypX encoding DNA SEQ ID NO:1.
XX
KW heterologous biological substance; Bacillus; cypX; red pigment; gene; ds.
OS Bacillus subtilis.
XX

Handwritten signature

Key Location/Qualifiers
CDS 1..1215
FT /*tag= a
FT /product= "cypX"
XX
PN WO2004011609-A2.
XX
PD 05-FEB-2004.
XX
PF 25-JUL-2003; 2003WO-US023398.
XX
PR 26-JUL-2002; 2002US-0398853P.
XX
PA (NOVO) NOVOZYMES BIOTECH INC.
XX
PI Tang M, Sloma A, Sternberg D, Behr R;
XX
WPI; 2004-143839/14.
DR P-PSDB; ADJ62039.
XX

Producing a heterologous biological substance comprises cultivating pigment-deficient mutants of Bacillus cell that directs synthesis of the heterologous biological substance and has a modification of the cypX and ymc genes.
Example 1; SEQ ID NO 1; 62pp; English.

The present invention describes a method for producing a heterologous biological substance comprising cultivating a mutant of a parent Bacillus cell in a medium suitable for the production of a heterologous biological

substance, and recovering the heterologous biological substance from the cultivation medium, where the mutant cell comprises a first nucleic acid sequence directing synthesis of the heterologous biological substance and a second nucleic acid sequence comprising a modification of at least one of the genes *cypX* and *yvmC*, which are involved in the production of a red pigment. The mutant cell is also deficient in the production of the red pigment compared to the parent *Bacillus* cell when cultivated under the same conditions. Also described: (1) a mutant of a parent *Bacillus* cell, comprising a first nucleic acid sequence directing synthesis of a heterologous biological substance and a second nucleic acid sequence comprising a modification of at least one of the genes *cypX* and *yvmC*, which are involved in the production of a red pigment, wherein the mutant cell is deficient in the production of the red pigment compared to the parent *Bacillus* cell when cultivated under the same conditions; and (2) obtaining a mutant of a parent *Bacillus* cell, comprising introducing into the parent *Bacillus* cell a first nucleic acid sequence directing synthesis of a heterologous biological substance and a second nucleic acid sequence comprising a modification of at least one of the genes *cypX* and *yvmC*, which are involved in the production of a red pigment, and identifying the mutant cell comprising the modified nucleic acid sequence, where the mutant cell is deficient in the production of the red pigment compared to the parent *Bacillus* cell when cultivated under the same conditions. The methods and compositions of the present invention are useful for producing biological substances, e.g. nucleic acid, polyamide, polyamine, polyol, polypeptide, or polysaccharide, in pigment-deficient mutants of *Bacillus* cells, and obtaining the pigment-deficient mutant cells. The present sequence encodes *Bacillus subtilis cypX*, which is used in the exemplification of the present invention.

Sequence 1215 BP; 323 A; 257 C; 300 G; 335 T; 0 U; 0 Other;

Query Match 100.0%; Score 1215; DB 12; Length 1215;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGCAATCGATTAAATGTTTGTAGTGTCTTCTGATCAATTTCAAAACAATCCATAT 60
DB 1 ATGAGCAATCGATTAAATGTTTGTAGTGTCTTCTGATCAATTTCAAAACAATCCATAT 60

QY 61 GCTTATTTTCAACATCGCGGAGGAAGATCCGGTTTCATTATGAAGATCGATAGACAGT 120
DB 61 GCTTATTTTCAACATCGCGGAGGAAGATCCGGTTTCATTATGAAGATCGATAGACAGT 120

QY 121 TATTTTATCAGCGGCTATCATGATGTCGGTATATCTTTCAGCATCCGATATCTTCAG 180
DB 121 TATTTTATCAGCGGCTATCATGATGTCGGTATATCTTTCAGCATCCGATATCTTCAG 180

QY 181 ACGAATCACTTGTGAGCGTCCGAAACAGATCATCGGAGGCCCTGTGCGCCCAATG 240
DB 181 ACGAATCACTTGTGAGCGTCCGAAACAGATCATCGGAGGCCCTGTGCGCCCAATG 240

QY 241 CATGGAAGAAACACTCTCCAAAGAAAGAAATGTTAGTGAGAAAGCTTTATCCGTTGACGA 300
DB 241 CATGGAAGAAACACTCTCCAAAGAAAGAAATGTTAGTGAGAAAGCTTTATCCGTTGACGA 300

QY 301 CTGGATCATCTCTCCATGATTAAACAAATGCAAAACTTGTAGGCGCTTATCTT 360
DB 301 CTGGATCATCTCTCCATGATTAAACAAATGCAAAACTTGTAGGCGCTTATCTT 360

QY 361 GAAAGAGGAAAGTAGTATCTGTCATGATTTTGGAAAGACGTTTCCGGTGTGCGTCACG 420
DB 361 GAAAGAGGAAAGTAGTATCTGTCATGATTTTGGAAAGACGTTTCCGGTGTGCGTCACG 420

QY 421 ATGGACATGCTCGGGCTGGATAAAGAGACCAATGAAATAATCTCTGAGTGGCACAGCGGA 480
DB 421 ATGGACATGCTCGGGCTGGATAAAGAGACCAATGAAATAATCTCTGAGTGGCACAGCGGA 480

QY 481 GTTCCGATTTTATCAGAGTATCTCTCAATCTCTGAGCGCGGCAATTCGTTATCG 540
DB 481 GTTCCGATTTTATCAGAGTATCTCTCAATCTCTGAGCGCGGCAATTCGTTATCG 540

QY 541 TGCAAGCAACAGCTTTCCCAATCTTGTATGCGCGGTCAATTAAGAAACGTCGCGTCAATCCG 600

DB 541 TGCAAGCAACAGCTTTCCCAATCTTGTATGCGCGTCAATTAAGAAACGTCGCGTCAATCCG 600
QY 601 GGATCAGATTTAAATTCGATCTCTATGTAATTAAGAGGATGCGGCTGTGCGAC 660
DB 601 GGATCAGATTTAAATTCGATCTCTATGTAATTAAGAGGATGCGGCTGTGCGAC 660
QY 661 AAGGATATCTCGCACTGATCTTAAATGCTGTTAGCGCAACCGGCTGATAAG 720
DB 661 AAGGATATCTCGCACTGATCTTAAATGCTGTTAGCGCAACCGGCTGATAAG 720
QY 721 ACGCTGGCACTGATGATCTACCAATTTGCTCAACAATCTCTGACGAGATGATGTTTG 780
DB 721 ACGCTGGCACTGATGATCTACCAATTTGCTCAACAATCTCTGACGAGATGATGTTTG 780
QY 781 GCTGACCGTTGCTAGTTCCGAGAGCCATTCGGGAGACATTCGGTTATAAACCGCGGTT 840
DB 781 GCTGACCGTTGCTAGTTCCGAGAGCCATTCGGGAGACATTCGGTTATAAACCGCGGTT 840
QY 841 CAGCTGATTTCCGCGCAGCTGTCCCAAGATACAGTGGTCCGCGGTATGAAATCAAAAA 900
DB 841 CAGCTGATTTCCGCGCAGCTGTCCCAAGATACAGTGGTCCGCGGTATGAAATCAAAAA 900
QY 901 GATACGATTTGTTTTTGTATGATCGGTGCGGTAACCGGACCCCTGAAGCATTTGAACAG 960
DB 901 GATACGATTTGTTTTTGTATGATCGGTGCGGTAACCGGACCCCTGAAGCATTTGAACAG 960
QY 961 CTTGACGTTGTTTAAATATTCATCGGGAAGATCTTGGTATCAAGAGCGCTTTTAGCGCGCC 1020
DB 961 CTTGACGTTGTTTAAATATTCATCGGGAAGATCTTGGTATCAAGAGCGCTTTTAGCGCGCC 1020
QY 1021 GCCCGGCATCTCGCTTTCCGATCCGCAATTCATACTGTGTAGGAGCAGCTTTTGCCAAA 1080
DB 1021 GCCCGGCATCTCGCTTTCCGATCCGCAATTCATACTGTGTAGGAGCAGCTTTTGCCAAA 1080
QY 1081 AACGAAATCGAAATTTGTAGTAATAATTGTCTGTAAGATCGGAATATCAGATTAGAG 1140
DB 1081 AACGAAATCGAAATTTGTAGTAATAATTGTCTGTAAGATCGGAATATCAGATTAGAG 1140
QY 1141 GAAATATTTTGTATGCTGAGTCCGGTCTGTATACACGGGACCTGTTTCACTTCTCGTT 1200
DB 1141 GAAATATTTTGTATGCTGAGTCCGGTCTGTATACACGGGACCTGTTTCACTTCTCGTT 1200
QY 1201 GCGTTTGAACGGGCA 1215
DB 1201 GCGTTTGAACGGGCA 1215

RESULT 2
ABK74890
ID ABK74890 standard; DNA; 1208 BP.
XX
AC ABK74890;
DT
DT 13-AUG-2002 (first entry)
XX
DE *Bacillus licheniformis* genomic sequence tag (GST) #2181.
XX
KW Differential gene expression; genomic sequenced tag; GST;
KW altered culture condition; environmental stress;
KW physiological provocation; ds.
XX
OS *Bacillus licheniformis*.
XX
FN WO200229113-A2.
XX
PD 11-APR-2002.
XX
PF 05-OCT-2001; 2001WO-US031437.
XX
PR 06-OCT-2000; 2000US-00680598.
PR 27-MAR-2001; 2001US-0279526P.
XX
PA (NOVO) NOVOZYMES BIOTECH INC.

(NOVO) NOVOZYMES AS.

Berka R, Clausen IG;

WPI; 2002-416584/44.

Monitoring differential expression of several genes in first *Bacillus* cell relative to expression of same genes in one or more second *Bacillus* cells, by using substrate containing *Bacillus* genomic sequenced tag array.

Claim 4; SEQ ID NO 2181; 200pp; English.

The invention describes a method of monitoring differential expression of genes in a first *Bacillus* cell relative to expression of the genes in other *Bacillus* cells, comprising hybridising labelled nucleic acid probes isolated from *Bacillus* cells to a substrate containing array of *Bacillus* genomic sequenced tags (GST), examining the array, and determining relative gene expression by an observed hybridisation reporter signal of a spot in the array. The method is useful for measuring the expression of genes in a first *Bacillus* cell relative to expression of the same genes in one or more second *Bacillus* cells. The method is useful for monitoring global expression of several genes from a *Bacillus* cell, discovering new genes, identifying possible functions of unknown open reading frames and monitoring gene copy number variation and stability. Monitoring changes in expression of genes may be used to provide a representation of the way in which *Bacillus* cells adapt to changes in culture conditions, environmental stress or other physiological provocation. Extensive follow-up characterisation is unnecessary, when one spot on an array equals one gene or one open reading frame, since sequence information is available. This sequence represents a genomic sequence tag (GST) used in the method of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 1208 BP; 318 A; 321 C; 307 G; 262 T; 0 U; 0 Other;

Query Match 50.8%; Score 617.2; DB 6; Length 1208;
Best Local Similarity 70.5%; Pred. No. 1.2e-186;
Matches 853; Conservative 0; Mismatches 353; Indels 4; Gaps 2;

```
QY 1 ATGAGCAATCGATTAAATGTTTGTAGTGTCTTTCTGATCAATTTCAAAACAAATCCATAT 60
DB 1 ATGATCAATCGTAAACAAATTCAGCGTGTGTGACAAATATCACGAAATCCGTAT 60
QY 61 GCTTATTTTCAACTCGGGAGGAGATCGGTTTCATATGACAGTCGATAGACGT 120
DB 61 CAGTATTTCTCATCCTTCGGGAATCTGATCCGCTCCATTATGAAGAAATCGCTGGACAGC 120
QY 121 TATTTTATCAGCGCTATCATGATGTCGCTATATCTTTCAGCATCCGGATATCTTCACG 180
DB 121 TATTTTATCAGCGCTATCAGGATGTCGCGGTGCTCCAGAAATCAGACGCTTTTACA 180
QY 181 ACGAAAT--CACTTGTGTAGCGTCCGAAACAGTCAATGCGAGGCCCTGTGTCGCCCAAA 238
DB 181 ACGATTTGCTGTGCTGCAACAGCGCGAGCCGCTCATGCGGACCTGTGTCGCCCAAA 240
QY 239 TGCATGGAAAAGAACACTCTGCGCAAAAGAAATTTGTAGTGAGAAAGCTTTATCGGTTGACG 298
DB 241 TGAAGGCAAAAGAGCACAGCGCAAAAGAGAAATCGTTTCTGCGCCGCTTTATCGGGGAAT 300
QY 299 CACTGATCATCTGCTCTCATTTGATTAACAAATCGAAGAACTGTTAGCGCTTATC 358
DB 301 CCGTTTATCACTTCACACCGCTCATCAAGAAATGCTCAAAAGCT--TTTGGCCCCGAC 358
QY 359 TTGAAAGAGGAAAAGTGTCTCGTCAATGATTTTGGAAAAGAGTGTGCGGTGTGCGTCA 418
DB 359 GTGGAAAGGCGGATCGACCTTGTCAATGATTTTCGCAAAACATTCGCGGTTTGGGTGA 418
QY 419 CGATGACATGCTCGGGCTGGATTAAGAGACCATGAAATAATCTCTGATGTGCGACAGCG 478
DB 419 CGATGACATTTTAGTTTGGACAGAAGACGACCACCAACCGGTGAGAAATCGGCACAGCG 478
```

```
QY 479 GAGTTCCGATTTTATCAGAGTATCTCTCAATCTCTGAACGCGGGGCACATTCGTTAT 538
DB 479 GCGTCGCGGATTTTCAATACAGTTTGAATCAGGCGCTGAGGACCGGGAGCATTCCTCA 538
QY 539 GGTGAGGAAACAGCTTTCCCAATACCTTGTATGATCCCGTCAATTAAGAAACCTCGGTCAATC 598
DB 539 AATGAGTGAACAGCTCGCTGAGTATTTGAATCCGATATTCGAGGAAAGCGGCAAAATC 598
QY 599 CCGGATCAGATTAATTTTCGATCTTATGATCTTCTGAATATGAAGGCATGCGGCTCGG 658
DB 599 CCGGACATGATTAATATCCATCTTTCGACATCCGAAATACGAGGAGTGGCGATGCTG 658
QY 659 ACAAGATATATCTCGACATGATTTTAATGTCTGTAGCCGCAACGGAACCGGCTGATA 718
DB 659 ACCGGACATACCGCGCTGATTTCTCAATATATTTGTTCCGCCACAGAACCGGGGACA 718
QY 719 AGACCTGGCAGTATGATCTACCAATTTGCTCAACAATCTCGACAGATGAATGATTTT 778
DB 719 AGACGCTGCATTAATGATTTATCAATTTGCTTCAATCCGATCAGATGAATGATGTC 778
QY 779 TGGCTGACCGTTTCGTTAGTTCCGAGAGCCATTCGCGAGACATTTGCGTTATAAACCCCGG 838
DB 779 TGAAGACCGCAGCTTCTCCCTCAGGCAATCGAGAGACGCTGCGCTACAGCCCGCGG 838
QY 839 TTCAGCTGATTCGCGGCGAGCTGTCCAAAGATACAGTGTGTCGCGGTATGGAAATCAAAA 898
DB 839 TGCAGCTCATCCGCGCCAGCTTTTCAAGAGCGCGAGATTGCGGAGTCGAGCTAAAG 898
QY 899 AGATACGATGTTTTTTTGTATGATCGGTGCGGTAAACCGGACCTGAAGCATTTGAAC 958
DB 899 AAGGAGCAGCTGTATTTTGCATGATAGCGCGGCAAAATCCGATCCTGAAGCGTTGAGG 958
QY 959 AGCCTGACGTGTTTAAATATTTTCATCGGGAAGATCTTGGTATCAAGAGCGCTTTTAGCGGCG 1018
DB 959 ATCCCGACAGTTCAACATTCACCGCAGCGACTGGAAGTCAAAAGCGCATTCAGCGGCG 1018
QY 1019 CCGCCCGGCATCTCGCTTTCCGATCCGCAATTCATAACTGTGTAGAGCAGCTTTTGCCA 1078
DB 1019 CAGCCAGGCATCTCGCTTTCCGTTTCCGTTTCCAGGCGTCCACAACTGTGTAGAGCGGGTTTGGCA 1078
QY 1079 AAAACGAAATCAAAATTTAGCTTAATATTTGCTGGAATGAAGATCGGATATCAGATTAG 1138
DB 1079 AGACCGAAATCAGCTTTGCGAAATATCGTCTTGTATCAGCTGAAAAATATCCGGCTGG 1138
QY 1139 AGAAGATTTTGTATGCTGAGTCCGCTCTGTATACGCGGACGACCTGTTTCACTTCTCG 1198
DB 1139 AGAAGACTTTCATTTTACCGGAGACAGGCTTTTACACGCGGACCCGCTCTCGCTCAACA 1198
QY 1199 TTGCGGTTTGA 1208
DB 1199 TCCGGTTTGA 1208
RESULT 3
ID ACF70115
XX ACF70115 standard; DNA; 1224 BP.
XX ACF70115;
XX ACF70115;
XX ACF70115;
DT 20-NOV-2003 (first entry)
XX
DE Photorhabdus luminescens nucleotide sequence #8582.
XX
KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW detection; food; gene expression; plant; animal; microorganism; toxin;
KW antibiotic; bioplastic; virulence factor; disease model; plague;
KW whooping cough; gene; ds.
XX
OS Photorhabdus luminescens.
XX
PN WO200294867-A2.
PD 28-NOV-2002.
```


83246 AATTCTATAAAATCCTTACGATTTTTCAGACATTTTTCATATAAACAAGATTAGTTTATT 83305
QY
101 ATGAAGAGTCGATAGACAGTTATTTTATCAGCGCTATCATGATGTCGCCGTATATCCTTC 160
Db
83306 TTGAACAATCCAAATAGCTATTTTATTTGGCAATATGAGGATGTTGACGCTATTTTAA 83365
QY
161 AGCATCCGGATATCTTCACGAGAAATCACTTGTGTGAGCGTCCGAAACCAAGTCATCGGAG 220
Db
83366 AATCTCTCA---ATTTCATATCAAGCCATTAACTGCACTTCCGAAACCGTATATGGGG 83422
QY
221 GCCCTGTGTCGCCCAATGCATGGAAGAACACACTCTCCCAAGAAAGAAATGTAGTCA 280
Db
83423 ATCGCGTCTTCTCAATGGAAGGTAAGAGATCGGTGTAAACGAAATTCATAATGC 83482
QY
281 GAAGCTTTATCGGTGACGCACTGGATC---ATCTGTCTCCATTGATTAACAAAATGCGAG 337
Db
83483 AAGGACTTTCAGAGATTATTTTATTCGCTACTACGAACCCATGATTCGAAGATTACTG 83542
QY
338 AAAAATGTTAGCGCTTATCTTGAAGAGGAAAGATGATCTCGTCAATGATTTTGGAA 397
Db
83543 AAGATCTCTTCAACCTTATATGGAAGAAAGAAATATAGATATTGTTAATGACTTTGGCC 83602
QY
398 AGACGTTTGGGTGTCGTCACGATGGACATGCTCGGGCTGGATTAAGAGACCATGAA 457
Db
83603 GCGATTATCCGCTATTAGTGACATTAAGATTATCTCGACTGCAAGTGATTAATCTCGT 83662
QY
458 AAATCTCTGAGTGGACACGCGAGTTGCGGATTTTATCACGAGTATCTCTCAATCTCTGT 517
Db
83663 ATATAGCTGATGACAGAGGTATCGGCAGTTTATTAACCCAGTTTGACCAACAGAAC 83722
QY
518 AAGCGCGGCAATTCGTTTATGGTGAGGAGCAAGCTTTCCAAATACTTGTATGTCGGGTCA 577
Db
83723 TAGAAAAAATGCACAGTCTTGAATGTAGTCAGAAATTAATTCGCTTACTAAGCCCTATA 83782
QY
578 TTAAGAACGTCGCTCAATCCGGATCAGATTAAATTTTCGATCTTATGCTCTGAT 637
Db
83783 TAGATCAGAGCGCGGTATATCCAGTAAAGCAATATATCTATTTCTG----- 83831
QY
638 ATGAAGGATGCGCTGTCGGAAGGATATCTCGACTGATCTTAAATGCTGTGTAG 697
Db
83832 -TCAGATACAGGATGTCATGAGTGAAATTTACAGACTGTGTTTAAACATCTCTATTAG 83890
QY
698 CGCAACGGAACCGGTGATAAGACGCTGGCACTGATGATCTACCAATTTGCTCAACAATC 757
Db
83891 CGGCTACAGAACCTGCTCAAGATATTAGCAATGATGCTTAACTAAATATCTAATC 83950
QY
758 CTGAGCAGATGAATGATGTTTGGCTGACCGTTGTTAGTTCGGAGAGCATTCGGAGA 817
Db
83951 CGAGTATGCTTGATGTAGTTCTCAAAAGATCGCAGCTTAGTTCGAGATGATTTGAAGAA 84010
QY
818 CATTCGGTTTAAACCGCGGTTTCAGCTGATTCGCGGCGAGCTGTCCCAAGATACAGTGG 877
Db
84011 CTTTACGTCGTGACATCCCGGTACAACTCATTCGAAGAGGCGAGTGAGACGTCACTA 84070
QY
878 TCGCGCGTATGGAATCAAAAATGATGATTTGTTTGTATGATCGGTGGGCTTAACC 937
Db
84071 TTTCTGATCGATATACCTAAAGGCGCTAGTATTTGTATGATGTTGGCAGCTAATC 84130
QY
938 GGGACCTCGAAGATTGAAAGCCCTGACGTGTTTAAATTCATCGGGAAGATCTTGGTA 997
Db
84131 GTGATCCATCGGTTTTTCAAAACCAAAATGAATTTGATCTATATCGAAGAAAAATACCA 84190
QY
998 TCAAGAGCGCTTTAGCGCGCGCGCGCTGCTGATTCGCGGCGAGCTGTCCCAAGATACAGTGG 1057
Db
84191 CTTTCTCCAGAAAGCAATTCGAAAAGACATTTAGCTTTTGGTGAGGTACTCATGCCT 84250
QY
1058 GTGTAGGAGCAGCTTTTGCACAAAACGAAATCGAAATGTAGTAAATATTTGCTGGATA 1117
Db
84251 GTGCTCGGCTGCAATTTTCACTCAGTCAGTTGGAAGTTTCAATCCATATATTTTGGATC 84310
QY
1118 AGATCGGGAATATCAGATTAGAGGAAGATTTTGTGTTATGCTAGTCCGCTGTATACAC 1177
Db
84311 TCCGTCATAATTTACGTTTGTGCTGATCACTATCATTATCAAGAAACAGCGGTGTATACAC 84370

QY 1178 GGGAGACCTGTTTCACTTCTCTGTTGCGTTTGA 1208
Db 84371 GAGGACCTTCTAAACTCCCTTTTGAGTTTGA 84401

RESULT 5

ACF67367_31
Continuation (32 of 57) of ACF67367 from base 3100001 (Photorhabdus luminescens nucleotic
WP Sequence split into 57 fragments LOCUS ACF67367 Accession ACF67367
WP Fragment Name Begin End
WP ACF67367_00 1 110000
WP ACF67367_01 100001 210000
WP ACF67367_02 200001 310000
WP ACF67367_03 300001 410000
WP ACF67367_04 400001 510000
WP ACF67367_05 500001 610000
WP ACF67367_06 600001 710000
WP ACF67367_07 700001 810000
WP ACF67367_08 800001 910000
WP ACF67367_09 900001 1010000
WP ACF67367_10 1000001 1110000
WP ACF67367_11 1100001 1210000
WP ACF67367_12 1200001 1310000
WP ACF67367_13 1300001 1410000
WP ACF67367_14 1400001 1510000
WP ACF67367_15 1500001 1610000
WP ACF67367_16 1600001 1710000
WP ACF67367_17 1700001 1810000
WP ACF67367_18 1800001 1910000
WP ACF67367_19 1900001 2010000
WP ACF67367_20 2000001 2110000
WP ACF67367_21 2100001 2210000
WP ACF67367_22 2200001 2310000
WP ACF67367_23 2300001 2410000
WP ACF67367_24 2400001 2510000
WP ACF67367_25 2500001 2610000
WP ACF67367_26 2600001 2710000
WP ACF67367_27 2700001 2810000
WP ACF67367_28 2800001 2910000
WP ACF67367_29 2900001 3010000
WP ACF67367_30 3000001 3110000
WP ACF67367_31 3100001 3210000
WP ACF67367_32 3200001 3310000
WP ACF67367_33 3300001 3410000
WP ACF67367_34 3400001 3510000
WP ACF67367_35 3500001 3610000
WP ACF67367_36 3600001 3710000
WP ACF67367_37 3700001 3810000
WP ACF67367_38 3800001 3910000
WP ACF67367_39 3900001 4010000
WP ACF67367_40 4000001 4110000
WP ACF67367_41 4100001 4210000
WP ACF67367_42 4200001 4310000
WP ACF67367_43 4300001 4410000
WP ACF67367_44 4400001 4510000
WP ACF67367_45 4500001 4610000
WP ACF67367_46 4600001 4710000
WP ACF67367_47 4700001 4810000
WP ACF67367_48 4800001 4910000
WP ACF67367_49 4900001 5010000
WP ACF67367_50 5000001 5110000
WP ACF67367_51 5100001 5210000
WP ACF67367_52 5200001 5310000
WP ACF67367_53 5300001 5410000
WP ACF67367_54 5400001 5510000
WP ACF67367_55 5500001 5610000
WP ACF67367_56 5600001 5648894

Query Match 20.4%; Score 247.4; DB 10; Length 110000;
Best Local Similarity 52.7%; Pred. No. 5.7e-67;
Matches 617; Conservative 0; Mismatches 536; Indels 18; Gaps 3;

QY 41 AATTTCAAAACAAATCCATATGCTTATTTTTCACAACTCGGGAGGAAGATCCGGTTCAAT 100
D8 88380 AATTTCATAAAAATCCTTACGATTTTACGACATTTTGCATAAACAGATTTAGTTTAT 88439
QY 101 ATGAAGAGTCGATAGACAGTATTTTATCAGCGGCTATCATGATGTCGCTATATCCTTC 160
D8 88440 TTGAACAATCCCAAAATAGCTATTTTATTTGGCAAAATATGAGGATGTGACGCTATTTAA 88499
QY 161 AGCATCCGATATCTTCACGACGAATCACTTGTGAGCGTCCGACCAAGTCATCGGAG 220
D8 88500 AATCTCTCA---ATTTTCAATACTAAGCCATTAACATGCACTTCCGAAACCGTTATGGGG 88556
QY 221 GCCCTGTGTCGCCCAAAATGCAATGAAAGAACACACTCTGCCAAAGAAAGAAATTTAGTGA 280
D8 88557 ATCGCGTCTTCTCTAATGGAGGTGAAGAGCATCGGTGAACGAAATTCATATGC 88616
QY 281 GAAGCTTTATCCGTGACGCACTGGATC---ATCTGTCTCCATTTGATTAACAAAAATGCA 337
D8 88617 AAGGACTTTTCAAGAGATTTATTTAATCGCTACTACGAACCCATGATTCGAAAGATTA 88676
QY 338 AAAACTTGTAGCGCTTATCTTGAAGAGGAAAGATGCTCGTCATGATTTTGGNA 397
D8 88677 AAGATCTCTTACAACCTTATATGGAAGAAAGGAATATAGATATTTGTAATGACTTTGGCC 88736
QY 398 AGACGTTTCGCGTGTGCGTCACGATCGATCGGCTGATAAAGAGAGACCATGAA 457
D8 88737 GCGATTATGCCGTTATAGTGACATTAAGATTTCTCGGACTGCCAAGTGAATACTATCG 88796
QY 458 AAATCTGAGTGGCAGCGGAGTTGCGGATTTTATATCACGAGTATCTCTCAATCTCCTG 517
D8 88797 ATATAGCTGAATGGCACAAGGCTATCGCCAGTTTATATCCAGTTTGACCAAAACAGAA 88856
QY 518 AAGCGGGGCAATTCGTTATGGTCAGAGCAACAGCTTTCCCAATACTTGATCGCGGTCA 577
D8 88857 TAGAAAAAATGCACAGTCTTGAATGTAGTCAGAAATTAATTCGCTTACTAAAGCCATA 88916
QY 578 TTAAGAGACGTCGCGTCAATCCGGATCAGATTTAATTTTCGATCTCTATCTCTGAT 637
D8 88917 TAGATCAGAGACGGGTAAATCCAGTAAGGACATTAATATCTATTTCTG----- 88965
QY 638 ATGAAGGCATGCGCTGTGTCGCAAGGATATATCTCGCACTGATTTCTTAATGTGCTGTAG 697
D8 88966 -TCAAGATACGCGATGTCATGAGTGAATTTACAGCACTGTGTTTAAACATCTTATTAG 89024
QY 698 CCGCAACGAAACGGCTGATAGACGCTGGACTGATGATCTACATTTGCTTCAACATC 757
D8 89025 CGGCTACAGAACTGTGCAAGAGATATTAGCAATGATGCTTAATCACTTAATCTAATC 89084
QY 758 CTGAGCAGATGAATGATGTTTGGCTGACCGTTCGTTAGTTCGAGAGCAATTCGGGAGA 817
D8 89085 CGAGTATGCTTATGATGTCTCAAGATCGCAGCTTAGTTCGAGATGATTTGAAGAA 89144
QY 818 CATTCGGTTATAAACCGCGGTTTCAGCTGATTTCCGCGCAGCTGTCCCAGATACAGTGG 877
D8 89145 CTTTACGCTGACATCCCGGTACAACTCATTTCCAGAGAGCGAGTGAAGAGCTCACTA 89204
QY 878 TCGCGGATGAATCAAAAAGATACGATTTGTTTGTATGATCGGTGCGGCTAAC 937
D8 89205 TTTCTGGTATGATATACCTAAAGGCGCTGTAGTATTTTGTATGATTTGCGCAGCTAATC 89264
QY 938 GGGACCTGAAGCTTTGAACAGCTGACGTGTTTAATTTCTATCTCGGAAGATCTTGTA 997
D8 89265 GTGATCCATCGGTTTTTCAAAACCAATGAATTTGATCTATATCGAAGAAAAATACCA 89324
QY 998 TCAAGAGCGCTTTTAGCGCGCGCGCGCATCTCGCTTTTCGGATCCGGCATTCATAACT 1057
D8 89325 CTTCTCCACAGAAAGCAATCGAAAGACATTTAGCTTTTGTGCGAGTACTCATGCT 89384
QY 1058 GTGTAGGACAGCTTTTGCACAAAACGAAATCGAAATTTGATGATTAATTTGCTGGATA 1117
D8 89385 GTGCTCGCGCTGATTTTCACTCAGTCAGTTGGAAGTTTCACTCAATATTTTGGATC 89444
QY 1118 AGATCGGAATATCAGATTAAGAGGAAGATTTTGTATGCTGAGTCCGCTCTGTATACAC 1177

D8 89445 TCCTGCATAATTTACGTTTGTCTGATCACTATTCATTAACAAGAACAGCGGTGTATACAC 89504
QY 1178 GCGGACCTGTTTCACTTCTCGTTGCGTTTGA 1208
D8 89505 GAGGACCTTCTAAACTCCTTTTGTAGTTTGA 89535
RESULT 6
AAA95667
ID AAA95667 standard; DNA; 1204 BP.
XX
AC AAA95667;
DT 14-FEB-2001 (first entry)
XX
DE Bacillus subtilis hydroxylating enzyme gene #2.
XX
KW Hydroxylated bicyclic compound; lactone; HMG-CoA; inhibitor; ds;
KW hydroxymethylglutaryl co-enzyme A reductase; hypocholesterolaemia.
XX
OS Bacillus subtilis.
XX
PN WO200044886-A1.
XX
PD 03-AUG-2000.
XX
PF 28-JAN-2000; 2000WO-JP000472.
XX
PR 29-JAN-1999; 99JP-00021707.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
XX Endo H, Yonetani Y, Mizoguchi H, Hashimoto S, Ozaki A;
XX WPI; 2000-548827/50.
DR P-PSDB; AAB15502.
XX
PT New protein derived from Bacillus genus microorganism useful for
PT producing a hydroxylated bicyclic compound e.g. HMG-CoA reductase
PT inhibitors.
XX
PS Claim 16; Page 90-94; 111pp; Japanese.
XX
CC This sequence represents the coding region for a novel Bacillus derived
CC protein having the activity of producing a hydroxylated bicyclic compound
CC or the corresponding lactone from a bicyclic compound or the
CC corresponding lactone. The protein is used for preparing hydroxylated
CC compounds useful as hydroxymethylglutaryl co-enzyme A (HMG-CoA) reductase
CC inhibitors and hypocholesterolaemic agents
XX
SQ Sequence 1204 BP; 333 A; 253 C; 316 G; 302 T; 0 U; 0 Other;
Query Match 6.7%; Score 82; DB 3; Length 1204;
Best Local Similarity 45.7%; Pred. No. 5.3e-15;
Matches 329; Conservative 0; Mismatches 385; Indels 6; Gaps 1;
QY 270 AATTGTAGTGAAAGCTTTATCGGTGAGCGCACTGATCATCTGTCTCCATTCATTAAACA 329
D8 289 AGTCGTGAACAAAGCCTTTTACTCCGCGCGATGAAGCAATGGGAACCGAGAATTCNAGA 348
QY 330 AAATGCAGAAACTTGTAGCGCCTTATCTTGAAGAGGGAAGATGATCTCTCAATCA 389
D8 349 AATCAACATGAACATGATTCAAAAATTTTCAGGGGCGCATGATTTGACCTTTTCACGA 408
QY 390 TTTTGGAAAGAGCTTTGCGGTGTCGCTCACGATGACATGCTCGGCTCGATAAAGAGA 449
D8 409 TTTTTCATACCGCTTCGGTTATTGTATATCTCAGCTGCTGGAGTGCCTTCAGCGCA 468
QY 450 CCATGAAAAAATCTCTGAGTGGCACAGCGGAGTTCGCGAGTTTATACAGAGTATCTCTCA 509
D8 469 TATGGAACAGTTTAAAGCATGTCTGTATCTCTGTCTAGTACACCGAAGGATAAAGATGA 528

QY 510 ATCTCTGTA-----GCGCGGCGACATTCGTTATGGTGCAGCGAACAGCTTCCCAATA 563
 Db |||||
 QY 529 AGAAGCTGAAAGAGCCCTTTTTCGGAAGACGAGATAAGTGTGAGGAAGAACTGGCCGCGTT 588
 Db |||||
 QY 564 CTTGATGCCGTCATTAAAGACGTCGCTCAATCCGGGATCAGATTTTAATTTTCGATCCT 623
 Db |||||
 QY 589 TTTTGGCGGCATCATAAGAAAGAAAGCGAAACAAACCGGAACAGGATATTATTCTATTTT 648
 Db |||||
 QY 624 ATGTACTTCTGAATATGAAGGCATGCGCTGTGCGCAAGGATATATCTCGCACTGATTC 683
 Db |||||
 QY 649 AGTGAAGCGGAAGAAACAGGCGAGAGCTGTCCGCTGAAGAGCTGATTCGCTGTGTCAC 708
 Db |||||
 QY 684 TAAATGCTGTAGTACCGCAACGGAACCGCGCTGATTAAGACCTGGCACTGATCTACCA 743
 Db |||||
 QY 709 GCTGCTGTGGTGGCGGGAATGAACACCACTCAACCTGATTTCAATGCGATGTACAG 768
 Db |||||
 QY 744 TTTGCTCAACAATCTGACGAGATGAATGATGTTTGGCTGACCGTTCGTTAGTTCCGAG 803
 Db |||||
 QY 769 CATATTAGAAACGCCAGGCTTTACAGGAACCTGCGCAGCCATCTGAACTGATGCTCA 828
 Db |||||
 QY 804 AGCCATTGCGGAGACATTCGTTTATAAACCGCGGTTCACTGATTCGCGGCGAGCTGTC 863
 Db |||||
 QY 829 GGCAGTGGAGGAAGCCCTTCGGTTTCAGAGCGCGCGCCCGTTTGAAGCGCATTGCCAA 888
 Db |||||
 QY 864 CCAAGATACAGTGTGCGCGGTATGGAATCAAAAAGATACGATTTTGTGTTATGAT 923
 Db |||||
 QY 889 GCGGATACGAGATCGCGGGGACCTGATTAAGAAGGATGATGTTGTTTGGCGTTGT 948
 Db |||||
 QY 924 CGTGGCGCTAACCGGACCTGAGCATTTGAACAGCTGACGCTGTTTAAATATCATCG 983
 Db |||||
 QY 949 GGCATCGGCAATCGTGATGAAGCAAGTTTGACAGCCGCATGTTTGTATTCGCGC 1008
 Db |||||

RESULT 7
 ACF06114
 ID ACF06114 standard; DNA; 1188 BP.
 XX
 AC ACF06114;
 DT 30-SEP-2003 (first entry)
 DE Bacterial P450 enzyme encoding DNA SEQ ID NO:11.
 XX
 KW Bacterial; P450; enzyme; catalysis; hydrolysis; epoxide; arene oxide;
 KW diol; alkene; chirality; thermotolerance; thermostability; gene; ds.
 OS Eubacteria.
 FH Key Location/Qualifiers
 FT CDS 1..1188
 FT /*tag= a
 FT /*product= "P450 enzyme"
 XX
 PN WO2003052050-A2.
 XX
 PD 26-JUN-2003.
 XX
 PF 05-AUG-2002; 2002WO-US024910.
 XX
 PR 03-AUG-2001; 2001US-0309497P.
 XX
 PA (DIVE-) DIVERSA CORP.
 XX
 PI Weiner D, Burke M, Hitchman T, Pujol C, Richardson T, Short J;
 XX
 DR WPI; 2003-541641/51.
 DR P-PSDB; ABR82124.
 XX
 PT Novel polypeptide having P450 enzyme activity and polynucleotides
 PT encoding the polypeptide, useful for catalyzing the hydrolysis of
 PT epoxides and arene oxides to their corresponding diols.
 XX
 PS Claim 18; Page 56; 365pp; English.

XX The present sequence encodes a bacterial P450 enzyme. P450 enzymes can be
 CC used to catalyse the hydrolysis of epoxides and arene oxides to their
 CC corresponding diols. P450 enzymes can also be used for hydrolysing an
 CC alkene for producing a compound of a desired chirality, and for
 CC increasing thermotolerance or thermostability of a P450 polypeptide
 XX
 SQ Sequence 1188 BP; 239 A; 371 C; 338 G; 240 T; 0 U; 0 Other;
 Query Match 6.7%; Score 81.4; DB 9; Length 1188;
 Best Local Similarity 52.2%; Pred. No. 8.1e-15;
 Matches 181; Conservative 0; Mismatches 166; Indels 0; Gaps 0;
 QY 640 GAAGCATGCGCTGTGCGACAAGGATATATCTCGCACTGATTCCTTAATGTGCTGTAGCC 699
 Db |||||
 QY 637 CATGCGCGCCCTCGACGATCAGCGCTGTGCGCAATATCTTTTCGCTGTGCTATCG 696
 Db |||||
 QY 700 GCAACGGAACGCGCTGATAAGACGCTGGCACATGATCTACCATTTGCTCAACATCCT 759
 Db |||||
 QY 697 GGTCTCGACACCGTGCCTGTCATCGGCGCGCAATCTATTCTGCGGAGCATCCG 756
 Db |||||
 QY 760 GAGCAGATGAATGATGTTTGGCTGACCGTTCGTTAGTTCCGAGAGCCATTTGGGAGACA 819
 Db |||||
 QY 757 GCGAGCTGGAGCGTGGCTCCGACCGCGCTGATTCGCGCGCTTCGCTGAGACC 816
 Db |||||
 QY 820 TTGCGTTATAAACCGCGGTTCACTGATTCGCGGCGAGCTGTCCCAAGATACAGTGGTC 879
 Db |||||
 QY 817 GTGCGCTACGATCAGCGCACCAATGCTACTGGCGCGACTGTTGCCATTGACACCGACAAA 876
 Db |||||
 QY 880 GCGGTATGGAATCAAAAAGATACGATTTTGTGATGATCGGTGCGGCTAACCGG 939
 Db |||||
 QY 877 TACGCAAGCCCATGAAAAAGGTCAAGCGGCTCTTTCATGATGTCGTCGCGCAACCGT 936
 Db |||||
 QY 940 GACCTGGAAGCATTTGAACAGCGCTGACGCTGTTTAAATATTTCATCGGA 986
 Db |||||
 QY 937 GACCGCTGGAATTCGACACACCCCGACAGTTCAATATATACCGGA 983
 Db |||||

RESULT 8
 AAA95666
 ID AAA95666 standard; DNA; 1191 BP.
 XX
 AC AAA95666;
 DT 14-FEB-2001 (first entry)
 DE Bacillus subtilis hydroxylating enzyme gene.
 XX
 KW Hydroxylated bicyclic compound; lactone; HMG-CoA; inhibitor; ds;
 KW hydroxymethylglutaryl co-enzyme A reductase; hypocholesterolaemia..
 OS Bacillus subtilis.
 XX
 PN WO200044886-A1.
 XX
 PD 03-AUG-2000.
 XX
 PF 28-JAN-2000; 2000WO-JP000472.
 XX
 PR 29-JAN-1999; 99JP-00021707.
 XX
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 XX
 PI Endo H, Yonetani Y, Mizoguchi H, Hashimoto S, Ozaki A;
 XX
 DR WPI; 2000-548827/50.
 DR P-PSDB; AAB15501.
 XX
 PT New protein derived from Bacillus genus microorganism useful for
 PT producing a hydroxylated bicyclic compound e.g. HMG-CoA reductase
 PT inhibitors.
 XX
 PS Claim 14; Page 64-67; 111pp; Japanese.

QY 804 AGCCATTGCGGAGACATTCGGTTATAAACCGCGGTTTCACTGATTCGCGCGCAGCTGC 863
DB 846 GGCAGTGGAGGAGCCTTGGCTTTCAGAGCGCGCGCCCGGTTTTCAGGCGCATTGCCAA 905
QY 864 CCAAGATACAGTGGTTCGGCGGTATGGAATCAAAAAGATACAGATGTTTTTGTATGAT 923
DB 906 GCGGATACGGAGATTCGGGGGCACCTGATTAAAGAGGTGATAGTTTTTGGCGCTTGT 965
QY 924 CGGTGCGGCTAACCGGACCTGAAGCATTTGAACAGCCTGACGTGTTTAAATATTCATCG 983
DB 966 GGCATCGGCAATCTGATGAAGCAAGTTTCAGACCGCCACATGTTTGATATCGCGG 1025

RESULT 10
AAA95669
ID AAA95669 standard; DNA; 1221 BP.
AC AAA95669;
XX
XX
DT 14-FEB-2001 (first entry)
XX
DE Bacillus subtilis hydroxylating enzyme gene #4.
XX
KW Hydroxylated bicyclic compound; lactone; HMG-CoA; inhibitor; ds;
KW hydroxymethylglutaryl co-enzyme A reductase; hypocholesterolaemia.
XX
OS Bacillus subtilis.
XX
PN WO200044886-A1.
XX
PD 03-AUG-2000.
PF 28-JAN-2000; 2000WO-JP000472.
XX
PR 29-JAN-1999; 99JP-00021707.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PI Endo H, Yonetani Y, Mizoguchi H, Hashimoto S, Ozaki A;
XX
DR WPI; 2000-548827/50.
DR P-PSDB; AAB15503.
XX
PT New protein derived from Bacillus genus microorganism useful for
PT producing a hydroxylated bicyclic compound e.g. HMG-CoA reductase
PT inhibitors.
XX
PS Claim 16; Page 101-104; 11pp; Japanese.
XX
CC This sequence represents the coding region for a novel Bacillus derived
CC protein having the activity of producing a hydroxylated bicyclic compound
CC or the corresponding lactone from a bicyclic compound or the
CC corresponding lactone. The protein is used for preparing hydroxylated
CC compounds useful as hydroxymethylglutaryl co-enzyme A (HMG-CoA) reductase
CC inhibitors and hypocholesterolaemic agents
XX
SQ Sequence 1221 BP; 334 A; 261 C; 322 G; 304 T; 0 U; 0 Other;

Query Match 6.4%; Score 77.2; DB 3; Length 1221;
Best Local Similarity 45.3%; Pred. No. 1.8e-13;
Matches 326; Conservative 0; Mismatches 388; Indels 6; Gaps 1;

QY 270 AATTGTAGTGAGAGCTTTATCGGTACGCATCTGATCATCTCTCCATTGATTAACA 329
DB 306 AGTCGTGAACAAGCCCTTTACTCCGCGCGATGAAGCAATGGGAACCGAGAAATCAAGA 365
QY 330 AAATGCAAGAACTTGTAGCGCTTATCTGAAAGAGGAAAGTATCTGTCATCAATGA 389
DB 366 ATACAGATGACTGATTCAAAATTTTCAGGGCGCAGTGTGACCTTGTTCACGA 425
QY 390 TTTTGGAAAGAGCTTTGCGGTGTCGATGAGCATGCTCGGCTGGATAAAGAGA 449

DB 426 TTATTCATACCCGCTTCGGTTATTTGTGATATCTGAGCTGCTGGAGTGCCCTTCAGCGCA 485
QY 450 CCATGAAAAAATCTCTGAGTGGCACAGCGGAGTTTCCGATTTTATACAGAGTATCTCTCA 509
DB 486 TATGGAACAGITTTAAAGCATGGTCTGATCTTCTGTAGTACACCGAAGGATAAAGTGA 545
QY 510 ATCTCTGAA-----GCGCGGCGCATTCGTTATGTTGTCAGCGAACAGCTTTCCCAATA 563
DB 546 AGAAGCTGAAAAAAGCCTTTTGGGAAGAACGAGATAGTGTGAGGAAGAACTGGCGCGCTT 605
QY 564 CTTGATGCGCGTCAATTAAGAAAGCTCGCGTCAATCCGGGATCAGATTTAAATTTTCATCT 623
DB 606 TTTTCCCGCATCATAGAGAAAGCGAAACAAACCGGAACAGGATATATTTCTATTTT 665
QY 624 ATGTACTTCTGAATATGAAGCATGCGCTGTCGGAACAAGGATATATCTGCACTGATTTCT 683
DB 666 AGTGAAGCGGAAGAAACAGCGGAGAGCTGTCCGGTGAAGAGCTGATTCCTGTTGTGCAC 725
QY 684 TAATGTCTGTAGCGCAACGGAACCGCTGATAGAGCTGCGCACTGGCACTGATCTACCA 743
DB 726 GCTGCTGCTGGTGGCGGAAATGAACCACTACAAACCTGATTTCAAAATGCGATGTTTCA 785
QY 744 TTTGCTCAACAATCTGAGCAGATGAATGATGTTTGGCTGACCGTTCGTTAGTTCGCGAG 803
DB 786 CATATTAGAACCGCAGCGGTTTACAGGAATCTGCGAGCCATCTGAACTGATGCCCA 845
QY 804 AGCCATTGCGGAGACATTCGGTTATAAACCGCGGTTTCACTGATTCGCGCGCAGCTGC 863
DB 846 GGCAGTGGAGGAAGCCTTGGCTTTCAGAGCGCGCGCGCTTTTGGAGCGCATTTGCCAA 905
QY 864 CCAAGATACAGTGTGCGCGGTATGGAATCAAAAAGATACAGTGTGTTTGTATGAT 923
DB 906 GCGGATACGGAGATCGGGGGGCACCTGATTAAAGAGGTGATAGCGTTTTCGCGCTTGT 965
QY 924 CGGTGCGCTAACCGGACCTTGAAGCATTTTGAACAGCCTGACGTGTTTAAATATTCATCG 983
DB 966 GGCATCGGCAATCTGATGAAGCAAGTTTTCAGACCGCAATGTTTGATATCGCGCG 1025

RESULT 11
ABK74899
ID ABK74899 standard; DNA; 966 BP.
XX
AC ABK74899;
XX
DT 13-AUG-2002 (first entry)
XX
DE Bacillus licheniformis genomic sequence tag (GST) #2190.
XX
KW Differential gene expression; genomic sequenced tag; GST;
KW altered culture condition; environmental stress;
XX physiological provocation; ds.
OS Bacillus licheniformis.
XX
PN WO200229113-A2.
XX
PD 11-APR-2002.
XX
PF 05-OCT-2001; 2001WO-US031437.
XX
PR 06-OCT-2000; 2000US-00680598.
PR 27-MAR-2001; 2001US-0279526P.
XX
PA (NOVO.) NOVOZYMES BIOTECH INC.
PA (NOVO.) NOVOZYMES AS.
XX
PI Berka R, Clausen IG;
XX
DR WPI; 2002-416684/44.
XX
PT Monitoring differential expression of several genes in first Bacillus
PT cell relative to expression of same genes in one or more second Bacillus

PT cells, by using substrate containing Bacillus genomic sequenced tag
PT array.
XX Claim 4; SEQ ID NO 2190; 200pp; English.

XX The invention describes a method of monitoring differential expression of
CC genes in a first Bacillus cell relative to expression of the genes in
CC other Bacillus cells, comprising hybridising labelled nucleic acid probes
CC isolated from Bacillus cells to a substrate containing array of Bacillus
CC genomic sequenced tags (GST), examining the array, and determining
CC relative gene expression by an observed hybridisation reporter signal of
CC a spot in the array. The method is useful for measuring the expression of
CC genes in a first Bacillus cell relative to expression of the same genes
CC in one or more second Bacillus cells. The method is useful for monitoring
CC global expression of several genes from a Bacillus cell, discovering new
CC genes, identifying possible functions of unknown open reading frames and
CC monitoring gene copy number variation and stability. Monitoring changes
CC in expression of genes may be used to provide a representation of the way
CC in which Bacillus cells adapt to changes in culture conditions,
CC environmental stress or other physiological provocation. Extensive follow
CC up characterisation is unnecessary, when one spot on an array equals one
CC gene or one open reading frame, since sequence information is available.
CC This sequence represents a genomic sequence tag (GST) used in the method
CC of the invention. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 966 BP; 252 A; 229 C; 269 G; 216 T; 0 U; 0 Other;

Query Match 6.0%; Score 72.4; DB 6; Length 966;
Best Local Similarity 45.3%; Pred. No. 5.6e-12;
Matches 345; Conservative 0; Mismatches 411; Indels 6; Gaps 2;

QY 228 GCTGCCCAATGCGATGGAAGAAGACACTCTGCCAAAAGAGAATGTAGTGAGAAGCTT 287
DB 17 GATGATCAGCATGATGCGCGAAGACACAGAGATCAGATCGTCAACAGGCTTT 76
QY 288 TATCGGTGACGACATGATCTGCTCCATTGATTAACAAATGCGAAGAACTTGT 347
DB 77 TACACGGCGTCATGAGGAGTGGAGCGCGCATTCGCACTGACGATCAATGCT 136
QY 348 AGCGCTTATCTTGAAGAGGAAAGTGTATCTCGTCAATGATTTTGGAAAGACGTTTCG 407
DB 137 CGCTGATGTCGCGCAGGAGAGATCGACTGTGACAGGATTTTCATATCCGTTGCC 196
QY 408 GGTGTGCTCAGATGGAATGCTCGGCTGGGATGAAGAGACATGAAGAAATCTCTGA 467
DB 197 CGTAATCGTCAATTCGGAATTCGCGGCTTCCTTGGTGTATAAGCATCATTTCAAAGA 256
QY 468 GTGGCAGACGGAGTTGCCGATTTTATCAGAG-TATCTCTCAATCTCTGAAGCGCGG 526
DB 257 ATGGTCCGACCTGTTGGTCAGCGCTCGGAAGAGCATCGCGCGAAGATGTGAACGAGTG 316
QY 527 CACATTCGTTATGG-----TGCAGGGAACAGGCTTTCCCAATCTTGATGCCCGTCATTAA 581
DB 317 GAAACATCAGGACCAAGGCGAAGAGAGCTGACCGCATCTTTGAAAAGATGTTGA 376
QY 582 AGAAGCTCGCTCAATTCGGGATCAGATTATTTTCGATCCTATGATCTTCGAAATATGA 641
DB 377 AGAGAAAAGACAAAACCTCGGAATGACCTGATTTTCGCTTTTGTATCAAAAGCGGCAAGA 436
QY 642 AGGCATGGCGCTGCGGACAAGGATATCTCGCACTGATCTTAATGTGCTGTAGCCGC 701
DB 437 AGGGACAAGCTGTCCCTCGATGAATGTTGCTTCAATTTGCAATCTGCTTTGATGGCGG 496
QY 702 AACGGAAACGGCTGATAGACGCTGGCACTGATGATCTACCAATTTGCTCAACATCTCTGA 761
DB 497 GAATGAACGACGACGAATTTAGTTTCAATCGGCTTACAGCATTTCTGAAACACCGGG 556
QY 762 GCAGATGAATGATGTTTGGCTGACGGTTCGTTAGTTTCGAGAGCCATTCGCGGACATTT 821
DB 557 CGTGTATGACGAGCTCGCGACGATCCTGAACTGATTTCCGCGAGGCGATCGAGGAGCGGT 616

QY 822 CGTTTATAAACCGCGGTTTCAGTTCGCGGCAGCTGTCCCAAGATACAGTGGTCCG 881
DB 617 CCGTTCCGGGCGCAGCGCGATGATTGTGGCTTCGTTAAACAGGATACCGAGATCAG 676
QY 882 CGGTATGGAATCAAAAAGATACAGATTGTTTTTGTATGATCGGTGCGGCTAACCGGA 941
DB 677 AGGAGTAAGCTTGA AAAAGAGAGGTGTGATCGCTTTCTTGTCTTCTGCGCAACCGTGA 736
QY 942 CCCTGACGATTTGAACAGCCTGAGCTGTTTAAATATTCATCG 983
DB 737 TGAACGAATTTGAGAGGCGCACGAATTTGATATTCACCG 778

RESULT 12

AAI99682_39/c
Continuation (40 of 45) of AAI99682 from base 3900001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682

Fragment Name	Begin	End
WP AAI99682_00	1	110000
WP AAI99682_01	100001	210000
WP AAI99682_02	200001	310000
WP AAI99682_03	300001	410000
WP AAI99682_04	400001	510000
WP AAI99682_05	500001	610000
WP AAI99682_06	600001	710000
WP AAI99682_07	700001	810000
WP AAI99682_08	800001	910000
WP AAI99682_09	900001	1010000
WP AAI99682_10	1000001	1110000
WP AAI99682_11	1100001	1210000
WP AAI99682_12	1200001	1310000
WP AAI99682_13	1300001	1410000
WP AAI99682_14	1400001	1510000
WP AAI99682_15	1500001	1610000
WP AAI99682_16	1600001	1710000
WP AAI99682_17	1700001	1810000
WP AAI99682_18	1800001	1910000
WP AAI99682_19	1900001	2010000
WP AAI99682_20	2000001	2110000
WP AAI99682_21	2100001	2210000
WP AAI99682_22	2200001	2310000
WP AAI99682_23	2300001	2410000
WP AAI99682_24	2400001	2510000
WP AAI99682_25	2500001	2610000
WP AAI99682_26	2600001	2710000
WP AAI99682_27	2700001	2810000
WP AAI99682_28	2800001	2910000
WP AAI99682_29	2900001	3010000
WP AAI99682_30	3000001	3110000
WP AAI99682_31	3100001	3210000
WP AAI99682_32	3200001	3310000
WP AAI99682_33	3300001	3410000
WP AAI99682_34	3400001	3510000
WP AAI99682_35	3500001	3610000
WP AAI99682_36	3600001	3710000
WP AAI99682_37	3700001	3810000
WP AAI99682_38	3800001	3910000
WP AAI99682_39	3900001	4010000
WP AAI99682_40	4000001	4110000
WP AAI99682_41	4100001	4210000
WP AAI99682_42	4200001	4310000
WP AAI99682_43	4300001	4410000
WP AAI99682_44	4400001	4411529

Query Match 5.6%; Score 67.8; DB 4; Length 110000;
Best Local Similarity 45.1%; Pred. No. 2.2e-09;
Matches 292; Conservative 0; Mismatches 352; Indels 3; Gaps 1;

QY 337 GAAACTTGTAGCCCTTATCTTGAAGAGGAAAGTATCTCGTCAATCATTTTGA 396
DB 55187 GACACCCCTGACGACCGCGTGTGCGAAGCGCGAGTGTGATCTTCTGCGGACCTGGCC 55128
QY 397 AAGACGTTTGGCGTGTGCGTCACGATGACATGCTCGGGCTGGATATAAAGAGACCATGAA 456


```
Db 55127 GCGCCGCTACCGATGCGGTGATCGCGCATGCTCGGGTGGCTCCAGAGCAGCGGAC 55068
Qy 457 AAAATCTCTAGTGGCAGACGGGAGTGGCGATTTTATCAGGATATCTCTCAATCTCT 516
Db 55067 ATGTTCTTGGGTGGTCCGACGATCTGGTGACATCTCTC--AGTTCGATGTCTCAA 55011
Qy 517 GAAGCGGGGACATTCGTTATGTTGTCGACGAGCAAGCTTTCCCAATCTTGTATGCGGGTC 576
Db 55010 GAGGATTTCCAGATCACCATGACGCTTTCGGGGCTTACAGCACTTACCGGGGCCACC 54951
Qy 577 ATTAAGAACGTCGCTCAATCCGGATCAGATTTAATTTCCGATCTATGATCTCTGAA 636
Db 54950 ATTGGGACGCGGAGCGGACCCACCGACGACCTGGTCAGCGTCTGGTGAGTTCGGAA 54891
Qy 637 FATGAAGCATGCGCTGTTCGACCAAGGATATCTCGCACTGATTTCTTAATGTGCTGTTA 696
Db 54890 GTTGACGGCGAGCGGTAAAGCAGCAGGAGTGGTCTATGAGACGCTGCTGATCTGATC 54831
Qy 697 GCGGCAACCGAACCGGCTGATAGACGCTGGCACTGATGATCTACATTTGCTCAACAAT 756
Db 54830 GCGGCGCAGCAGACCAACCGCGGATACCTTCAGCGGTGTACCGAGCAGCTGCTGGCAAC 54771
Qy 757 CTTGACGATGATGATGTTTGGCTGACCGTTCGTTAGTTCCGAGAGCCATTCGGAG 816
Db 54770 CTTGACGATGATGATGTTTGGCTGACCGTTCGTTAGTTCCGAGAGCCATTCGGAG 54711
Qy 817 ACATTCGCTTATAACCGCGGTTCAGCTGATTCGCGCGCAGCTGTCCCAAGATACAGTG 876
Db 54710 ATGCTAGTTGACCGCGCGGTAAAGACATGTGCGGGGTGTGACCGCGGATACCGAG 54651
Qy 877 GTGCGGCTATGGAATCAAAAAGATACGATTTGTTTGTATGATGATGCTGCGGCTAAC 936
Db 54650 TTTTCAGCGCAGCGGCTGTGTGCGCGGAGAGATGATGCTCTTTCAGTTCGGCGAAC 54591
Qy 937 CCGGACCTTGAGCATTTGAAAGCTGACGTTTAAATATTCATCG 983
Db 54590 TTTTCAGCGCGGCTTTCTGTGAACCGGAAAAGTTGATGTTTCAGCG 54544
```

RESULT 13

AAI199683_39/c
Continuation (40 of 44) of AAI199683 from base 3900001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments LOCUS AAI199683 Accession AAI199683

Fragment Name	Begin	End
WP AAI199683_00	1	110000
WP AAI199683_01	100001	210000
WP AAI199683_02	200001	310000
WP AAI199683_03	300001	410000
WP AAI199683_04	400001	510000
WP AAI199683_05	500001	610000
WP AAI199683_06	600001	710000
WP AAI199683_07	700001	810000
WP AAI199683_08	800001	910000
WP AAI199683_09	900001	1010000
WP AAI199683_10	1000001	1110000
WP AAI199683_11	1100001	1210000
WP AAI199683_12	1200001	1310000
WP AAI199683_13	1300001	1410000
WP AAI199683_14	1400001	1510000
WP AAI199683_15	1500001	1610000
WP AAI199683_16	1600001	1710000
WP AAI199683_17	1700001	1810000
WP AAI199683_18	1800001	1910000
WP AAI199683_19	1900001	2010000
WP AAI199683_20	2000001	2110000
WP AAI199683_21	2100001	2210000
WP AAI199683_22	2200001	2310000
WP AAI199683_23	2300001	2410000
WP AAI199683_24	2400001	2510000
WP AAI199683_25	2500001	2610000
WP AAI199683_26	2600001	2710000
WP AAI199683_27	2700001	2810000

Query Match 5.6%; Score 67.8; DB 4; Length 110000;
Best Local Similarity 45.1%; Pred. No. 2.2e-09;
Matches 292; Conservative 0; Mismatches 352; Indels 3; Gaps 1;

Qy 337 GAAATCTTGTAGCGCTTATCTTGAAGAGGAAAGATGATCTCGTCAATGATTTTGA 396
Db 47917 GACACCTGATCGACGCGGTGTCGAAACGCGCGAGTGACCTTCTGCGGACCTGGCC 47858
Qy 397 AAGAGCTTTGCGGTGTGCGTACAGATGACATGCTCGGGCTGGATAAAGAGACCAATGA 456
Db 47857 GCGCGCTACCGATGCGCGGTGATCGCGACATGCTCGGGTGGTCCAGAGCAGCGGAC 47798
Qy 457 AAAATCTCTGAGTGACAGCGGAGTTCGCGATTTTATCAGGATATCTCTCAATCTCCT 516
Db 47797 ATGTTCTTTCGCGTGTGCGACGATCTGTTGACATTCCTC---AGTTCGATGTCTCAA 47741
Qy 517 GAAGCGCGGACATTCGTTATGTTGTCGACGCAACAGCTTTTCGCAATCTTGTATGCGGTC 576
Db 47740 GAGGATTTCCAGATCACCATGACGCGCTTCGCGGCTACACGACTTCACCGGGCCACC 47681
Qy 577 ATTAAGAACGTCGCGTCAATCCGGGATCAGATTTAAATTTTCGATCTTATGTTCTTCTGAA 636
Db 47680 ATTGCGGACGCGGAGCGGACCCACCGACGACCTGGTTCAGCGTGTGTTGAGTTCGAA 47621
Qy 637 TATGAAGCATGCGCTGTTCGACAGGATATATCGCACTGATCTTCTTATGTTGCTGTTA 696
Db 47620 GTTGAAGCGCGGCTTAAAGCAGCAGCGAGTGGTCTATGAGACGCTGCTGATCTGATC 47561
Qy 697 GCGCAACGGAACCGCTGATAGACGCTGGCACTGATCTTACCATTTGCTCAACAAT 756
Db 47560 GCGCGGACGAGACCAACGCGGCTATCTTGAAGCGGTGGTACCGAGCAGCTGCTGCGCAAC 47501
Qy 757 CTTGAGCAGATGAATGATGTTTGGCTGACCGTTCGTTAGTTCCGAGAGCCATTGCGGAG 816
Db 47500 CTTGACGAGTGGACCTGCTGTCAGCGGACCGCTGCTGCTCGCGGGCCATCGAGGAG 47441
Qy 817 ACATTCGCTTATAACCGCGGTTTTCAGTGATTCGCGGACGCTGCTCCCAAGATACAGTG 876
Db 47440 ATGCTACGTTGACGCGCGGCTTAAAGAACATGTCGCGGTTTGAACCGGATACCGAG 47381
Qy 877 GTCGCGGTATGGAATCAAAAAGATACGATTTGTTTGTATGATGATCGTGGCGCTAAC 936
Db 47380 TTTTCAGGACGCGGCTTGTGTCGCGGAGAGATGATGCTGCTCTTTCAGTTCGCGCAAC 47321
Qy 937 CCGGACCTTGAGCATTTTGAACAGCGCTGACGCTGTTTAAATATTCATCG 983
Db 47320 TTTTCAGGAGCGGCTTTTCTGTGAACCGGAAAAGTTGATGTTTCAGCG 47274

RESULT 14

ABK74891
ID ABK74891 standard; DNA; 1213 BP.
XX
AC ABK74891;
XX
DT 13-AUG-2002 (first entry)

13-AUG-2002

13-AUG-2002

XX Bacillus licheniformis genomic sequence tag (GST) #2182.
DE
XX Differential gene expression; genomic sequenced tag; GST;
KW altered culture condition; environmental stress;
KW physiological provocation; ds.
XX
XX Bacillus licheniformis.
XX
PN WO200229113-A2.
XX
XX 11-APR-2002.
XX
XX 05-OCT-2001; 2001WO-US031437.
XX
XX 06-OCT-2000; 2000US-00680598.
XX 27-MAR-2001; 2001US-0279526P.
XX
XX (NOVO) NOVOZYMES BIOTECH INC.
PA (NOVO) NOVOZYMES AS.
XX
XX Berka R, Clausen IG;
PI
XX WPI; 2002-416684/44.
DR
XX Monitoring differential expression of several genes in first Bacillus
PT cell relative to expression of same genes in one or more second Bacillus
PT cells, by using substrate containing Bacillus genomic sequenced tag
PT array.
XX
XX Claim 4; SEQ ID NO 2182; 200pp; English.
PS
XX The invention describes a method of monitoring differential expression of
CC genes in a first Bacillus cell relative to expression of the genes in
CC other Bacillus cells, comprising hybridising labelled nucleic acid probes
CC isolated from Bacillus cells to a substrate containing array of Bacillus
CC genomic sequenced tags (GST), examining the array, and determining
CC relative gene expression by an observed hybridisation reporter signal of
CC a spot in the array. The method is useful for measuring the expression of
CC genes in a first Bacillus cell relative to expression of the same genes
CC in one or more second Bacillus cells. The method is useful for monitoring
CC global expression of several genes from a Bacillus cell, discovering new
CC genes, identifying possible functions of unknown open reading frames and
CC monitoring gene copy number variation and stability. Monitoring changes
CC in expression of genes may be used to provide a representation of the way
CC in which Bacillus cells adapt to changes in culture conditions,
CC environmental stress or other physiological provocation. Extensive follow
CC -up characterisation is unnecessary, when one spot on an array equals one
CC gene or one open reading frame, since sequence information is available.
CC This sequence represents a genomic sequence tag (GST) used in the method
CC of the invention. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1213 BP; 340 A; 286 C; 287 G; 300 T; 0 U; 0 Other;

Query Match 5.3%; Score 64.8; DB 6; Length 1213;
Best Local Similarity 47.0%; Pred. No. 1.7e-09;
Matches 201; Conservative 0; Mismatches 227; Indels 0; Gaps 0;

QY 560 AATACTTGTATCGCGTCAATTAAGAAGCTGCGCTCAATCGGGATCAGATTTAATTGCA 619
DB 589 ATTATTTTAAACATNTTCAAGCAGCGGCAAAAAGAGGCAAAAGACGCTGATTTCCG 648
QY 620 TCCTATGTACTTCTGAATATGAAGGATCGCGTTCGGAACAAGGATATCTCGCACTGA 679
DB 649 TTTTACTCGCGGGGAAGTTGACGGCAATCGCTGACAGAAGAAGAACTGCTTCATTTT 708
QY 680 TTCCTAATGTCTGTAGCGCAACCGACCGCTGATAGACGCTGGCACTGATGATCT 739
DB 709 GCATCATCTTTTGGTCGCGAGGCAATGAGACGCAACCAACTTGATCGCAAAACGGTCC 768
QY 740 ACCATTGTCTCAACAATCCTGACAGATGAATGATGTTTTGGCTGACCGTTGCTTAGTTC 799

RESULT 15
AAI99682_08/c
Continuation (9 of 45) of AAI99682 from base 800001 (Mycobacterium tuberculosis strain H
WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682

Db	Fragment Name	Begin	End
769	AAI99682_00	1	110000
QY	AAI99682_01	100001	210000
DB	AAI99682_02	200001	310000
QY	AAI99682_03	300001	410000
DB	AAI99682_04	400001	510000
QY	AAI99682_05	500001	610000
DB	AAI99682_06	600001	710000
QY	AAI99682_07	700001	810000
DB	AAI99682_08	800001	910000
QY	AAI99682_09	900001	1010000
DB	AAI99682_10	1000001	1110000
QY	AAI99682_11	1100001	1210000
DB	AAI99682_12	1200001	1310000
QY	AAI99682_13	1300001	1410000
DB	AAI99682_14	1400001	1510000
QY	AAI99682_15	1500001	1610000
DB	AAI99682_16	1600001	1710000
QY	AAI99682_17	1700001	1810000
DB	AAI99682_18	1800001	1910000
QY	AAI99682_19	1900001	2010000
DB	AAI99682_20	2000001	2110000
QY	AAI99682_21	2100001	2210000
DB	AAI99682_22	2200001	2310000
QY	AAI99682_23	2300001	2410000
DB	AAI99682_24	2400001	2510000
QY	AAI99682_25	2500001	2610000
DB	AAI99682_26	2600001	2710000
QY	AAI99682_27	2700001	2810000
DB	AAI99682_28	2800001	2910000
QY	AAI99682_29	2900001	3010000
DB	AAI99682_30	3000001	3110000
QY	AAI99682_31	3100001	3210000
DB	AAI99682_32	3200001	3310000
QY	AAI99682_33	3300001	3410000
DB	AAI99682_34	3400001	3510000
QY	AAI99682_35	3500001	3610000
DB	AAI99682_36	3600001	3710000
QY	AAI99682_37	3700001	3810000
DB	AAI99682_38	3800001	3910000
QY	AAI99682_39	3900001	4010000
DB	AAI99682_40	4000001	4110000
QY	AAI99682_41	4100001	4210000
DB	AAI99682_42	4200001	4310000
QY	AAI99682_43	4300001	4410000
DB	AAI99682_44	4400001	441529

Query Match 5.1%; Score 61.8; DB 4; Length 110000;
Best Local Similarity 46.8%; Pred. No. 1.9e-07;

	Matches 195;	Conservative	0;	Mismatches 222;	Indels	0;	Gaps	0;
QY 608	ATTAAATTTTCGATCTTACTTCTGAATATGAAGGCATGGCGCTGTTCGGACAAGGATA	667						
Db 59445	ATCTGACGTCGGCACTGTCTGGCAGCCGAGCTCGACGGCGACCGGCTTTCCGACCAAGAAA	59386						
QY 668	TACTCGCACTGATCTTAATGTGCTGTAGCCGCAACGGACCGGCTGATAAGACGCTGG	727						
Db 59385	TCATGGCGTTCCTGTCTTCATGGTGATCGCCGGCAACGAGACCAACCAAGCTACTGG	59326						
QY 728	CACTGATGATCTACCATTTTGTCTCAACAATCTCTGAGCAGATGAATGATGTTTGGCTGACC	787						
Db 59325	CCAATGCCGTCTACTGGGCGGCCCAACCACTGGCCGAGCTGGCCCGGTATTGCGCGACC	59265						
QY 788	GTTCTGTAGTTCGAGAGCCATTCGGGAGACATTCGGTTATATAACCCCGGTTCACTGA	847						
Db 59265	ACTCCCGGATTCGATGTGGGTGGAGGAAACCTGCGCTACGACACGTCCAGCCAGATTC	59206						
QY 848	TTCCCGCGGAGCTGTCCCAAGATACAGTGTGGCGGTATGGAAATCAAAAAGATACGA	907						
Db 59205	TGGCCCGCACCGTCCGCGCACGATCTCACGTTGTACGACACCAACGATCCCGAGGTGAGG	59146						
QY 908	TTGTTTTTTTGTATGATCGGTGGGCTAACCGGACCTGAAGCATTTGAACAGCCTGACG	967						
Db 59145	TGTTGCTGTCTACCGGATCGGCCAACCGTGACACCGGGTGTTCGAGACCCCGAGC	59086						
QY 968	TGTTTAATATTCATCGGGAAGATCTTTGGTATCAAGAGCGCTTTTAGCGGCGCGCCC	1024						
Db 59085	ACTATCGCATCGCGCGGAAATCGGCTGCAAACTAGTCACTTTCGCGAGCGGTGCC	59029						

Search completed: December 10, 2004, 13:46:19
Job time : 681.942 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2004, 16:12:37 ; Search time 721.715 Seconds
(without alignments)
9281.958 Million cell updates/sec

Title: US-10-627-124-1
Perfect score: 1215
Sequence: 1 atgagccaatcgattaaatt.....tcgttcgcttgacggggca 1215

Scoring table: IDENTITY_NUC
Gapop 10_0 , Gapext 1.0

Searched: 4085731 seqs, 2756760397 residues

Total number of hits satisfying chosen parameters: 8171462

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA.*

- 1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq.*
- 3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq.*
- 5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq.*
- 7: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq.*
- 8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/prodata/2/pubpna/US09_PUBCOMB.seq.*
- 10: /cgn2_6/prodata/2/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/prodata/2/pubpna/US09_PUBCOMB.seq.*
- 12: /cgn2_6/prodata/2/pubpna/US09_PUBCOMB.seq.*
- 13: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq.*
- 14: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq.*
- 15: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq.*
- 16: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq.*
- 17: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq.*
- 18: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq.*
- 19: /cgn2_6/prodata/2/pubpna/US11_NEW_PUB.seq.*
- 20: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq.*
- 21: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1215	100.0	1215	16	US-10-627-124-1
2	617.2	50.8	1208	9	US-09-974-300-2181
3	113.6	9.3	2731748	17	US-10-297-465A-1
4	89.4	7.4	2731748	17	US-10-297-465A-1
5	81.4	6.7	1188	15	US-10-214-446-11
6	72.4	6.0	966	9	US-09-974-300-2190
7	67.6	5.6	1215	15	US-10-156-761-1
8	67.6	5.6	9025608	15	US-10-156-761-4523
9	64.8	5.3	1213	9	US-09-974-300-2182
10	55.6	4.6	1200	16	US-10-282-122A-25888
11	55.6	4.6	84428	15	US-10-229-148B-1
12	55.2	4.5	1233	15	US-10-205-032-7

c	13	55.2	4.5	60196	15	US-10-205-032-1	Sequence 1, Appli
	14	55	4.5	1155	9	US-09-974-300-2185	Sequence 2185, Ap
	15	53.6	4.4	1215	15	US-10-214-446-49	Sequence 49, Appl
	16	53.4	4.4	1248	15	US-10-156-761-7447	Sequence 7447, Ap
c	17	50	4.1	17596	18	US-10-611-442-2	Sequence 2, Appli
	18	49.6	4.1	1191	15	US-10-214-446-13	Sequence 13, Appl
	19	48.2	4.0	9024	18	US-10-611-442-1	Sequence 1, Appli
	20	46.6	3.8	1290	9	US-09-738-626-617	Sequence 617, App
c	21	46.6	3.8	1407	17	US-10-781-014-779	Sequence 779, App
	22	46.6	3.8	3309400	9	US-09-738-626-1	Sequence 1, Appli
	23	46.2	3.8	1248	15	US-10-214-446-55	Sequence 55, Appl
c	24	44.8	3.7	82993	15	US-10-080-170-645	Sequence 645, App
	25	44.8	3.7	82993	17	US-10-080-170-645	Sequence 645, App
c	26	44.6	3.7	82993	18	US-10-468-356-645	Sequence 39, Appl
	27	44.6	3.7	1215	15	US-10-214-446-39	Sequence 94, Appl
	28	43.6	3.6	1293	14	US-10-145-415-94	Sequence 21, Appl
	29	43	3.5	1209	9	US-09-861-289-21	Sequence 21, Appl
	30	43	3.5	1209	9	US-09-860-846-21	Sequence 21, Appl
	31	43	3.5	1209	10	US-09-988-384B-21	Sequence 21, Appl
	32	43	3.5	1209	10	US-09-836-821-21	Sequence 21, Appl
	33	43	3.5	1209	15	US-10-271-889-21	Sequence 21, Appl
	34	43	3.5	1209	17	US-10-398-605-21	Sequence 21, Appl
	35	43	3.5	12441	10	US-09-988-384B-3	Sequence 3, Appli
	36	43	3.5	13613	9	US-09-861-289-3	Sequence 3, Appli
	37	43	3.5	13613	9	US-09-860-846-3	Sequence 3, Appli
	38	43	3.5	13613	10	US-09-836-821-3	Sequence 3, Appli
	39	43	3.5	13613	15	US-10-271-889-46	Sequence 46, Appl
	40	43	3.5	13613	17	US-10-398-605-3	Sequence 3, Appli
	41	43	3.5	38506	10	US-09-793-708-19	Sequence 19, Appl
	42	43	3.5	38506	15	US-10-201-385-1	Sequence 1, Appli
	43	43	3.5	38506	15	US-10-160-539-19	Sequence 19, Appl
	44	42.6	3.5	1179	15	US-10-156-761-1975	Sequence 1975, Ap
c	45	42.6	3.5	9025608	15	US-10-156-761-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-10-627-124-1
; Sequence 1, Application US/10627124
; Publication No. US2004009694AI
; GENERAL INFORMATION:
; APPLICANT: NOVOZYMES BIOTECH, INC.
; TITLE OF INVENTION: Methods For Producing Biological Substances In Pigment-Deficient
; TITLE OF INVENTION: Mutants Of Bacillus Cells
; FILE REFERENCE: 10302.200-WO
; CURRENT APPLICATION NUMBER: US/10/627,124
; PRIOR FILING DATE: 2003-07-25
; PRIOR APPLICATION NUMBER: US 60/398,853
; PRIOR FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 1215
; TYPE: DNA
; ORGANISM: Bacillus subtilis
US-10-627-124-1

Query Match	100.0%	Score 1215;	DB 16;	Length 1215;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 1215;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
Qy	1	ATGAGCCCAATCGATTAAATTTGTTTGTCTGTCATTTTCACAAACATCCATAT	60	
Db	1	ATGAGCCCAATCGATTAAATTTGTTTGTCTGTCATTTTCACAAACATCCATAT	60	
Qy	61	GCTTATTTTTCACAACTCGCGGAGGAGATCCGTTTCATTATGAGAGTCGATAGACAGT	120	
Db	61	GCTTATTTTTCACAACTCGCGGAGGAGATCCGTTTCATTATGAGAGTCGATAGACAGT	120	
Qy	121	TATTTTATCAGCGGCTATCATGATTCGCTATATCTTTTACGATCCGATATCTTCACG	180	
Db	121	TATTTTATCAGCGGCTATCATGATTCGCTATATCTTTTACGATCCGATATCTTCACG	180	

Applicant

Db 121 TATTTTATCAGCGGCTATCATGATGTCGCTATATCTCTTTCAGCATCGGATATCTTTCAG 180
Qy 181 ACGAATCACTTTGTTGAGGTCGCCAACCAGTCATGCGAGGCCCTGTGTCGCCCAATG 240
Db 181 ACGAATCACTTTGTTGAGGTCGCCAACCAGTCATGCGAGGCCCTGTGTCGCCCAATG 240
Qy 241 CATGAAAAGAACACTCTGCCAAAAGAAATTTAGTAGAGAGCTTTTATCGGTGACGCA 300
Db 241 CATGAAAAGAACACTCTGCCAAAAGAAATTTAGTAGAGAGCTTTTATCGGTGACGCA 300
Qy 301 CTGGATCATCTCTCCATTTGATTAACAAAATGAGAAAATTTGTTAGCGCTTATCTT 360
Db 301 CTGGATCATCTCTCCATTTGATTAACAAAATGAGAAAATTTGTTAGCGCTTATCTT 360
Qy 361 GAAAGGGGAAAGTATCTCTCAATGATTTTGGAAAGACCTTTTCCGCTGTCAG 420
Db 361 GAAAGGGGAAAGTATCTCTCAATGATTTTGGAAAGACCTTTTCCGCTGTCAG 420
Qy 421 ATGGACATGCTCGGCTGGATTAAGAGACCATGAAATAATCTCTAGTGGCACAGCGGA 480
Db 421 ATGGACATGCTCGGCTGGATTAAGAGACCATGAAATAATCTCTAGTGGCACAGCGGA 480
Qy 481 GTTGGCGATTTTATCAGAGTATCTCTCAATCTCTGAAGCGCGGCAATTCGTTATGG 540
Db 481 GTTGGCGATTTTATCAGAGTATCTCTCAATCTCTGAAGCGCGGCAATTCGTTATGG 540
Qy 541 TCGAGGAAACAGCTTTCCCAATCTGATGCGCGGTCAATTAAGAGCGTCCGCTCAATCG 600
Db 541 TCGAGGAAACAGCTTTCCCAATCTGATGCGCGGTCAATTAAGAGCGTCCGCTCAATCG 600
Qy 601 GGATCAGATTTAATTTTCGATCTCTGATCTCTGAATATGAAGGATGCGCTGCGGAC 660
Db 601 GGATCAGATTTAATTTTCGATCTCTGATCTCTGAATATGAAGGATGCGCTGCGGAC 660
Qy 661 AAGGATATCTCGACTGATTTCTTAATGTGCTTTAGCGGCAACCGGCTGATAG 720
Db 661 AAGGATATCTCGACTGATTTCTTAATGTGCTTTAGCGGCAACCGGCTGATAG 720
Qy 721 ACGTGGCACTGATGATCTACCATTTGCTCAACATCTCGAGCAGATGATGTTTG 780
Db 721 ACGTGGCACTGATGATCTACCATTTGCTCAACATCTCGAGCAGATGATGTTTG 780
Qy 781 GCTGACCGTTCTGTTAGTCCGAGAGCAATGCGGAGACATTCGTTTATAAACCGCGGTT 840
Db 781 GCTGACCGTTCTGTTAGTCCGAGAGCAATGCGGAGACATTCGTTTATAAACCGCGGTT 840
Qy 841 CAGCTGATTTCCGCGGAGCTGTCCCAAGATACAGTGTGCGCGGTATGGAATCAAAAA 900
Db 841 CAGCTGATTTCCGCGGAGCTGTCCCAAGATACAGTGTGCGCGGTATGGAATCAAAAA 900
Qy 901 GATACGATTTGTTTGTATGATCGTGGCTAACCGGACCTTGAAGCATTTGAACAG 960
Db 901 GATACGATTTGTTTGTATGATCGTGGCTAACCGGACCTTGAAGCATTTGAACAG 960
Qy 961 CTTGACGCTGTTTAAATTCATCGGGAAGATCTTGGTATCAAGAGCGCTTTTAGCGCGC 1020
Db 961 CTTGACGCTGTTTAAATTCATCGGGAAGATCTTGGTATCAAGAGCGCTTTTAGCGCGC 1020
Qy 1021 GCCCGGCACTCTCGCTTTCCGATCCGGCAATTCATACTGTGTAGGAGCAGCTTTTGCCAA 1080
Db 1021 GCCCGGCACTCTCGCTTTCCGATCCGGCAATTCATACTGTGTAGGAGCAGCTTTTGCCAA 1080
Qy 1081 AACGAAATCGAAATTTAGCTAATATTGCTGGATTAAGATCGGGAATATCAGATTAGAG 1140
Db 1081 AACGAAATCGAAATTTAGCTAATATTGCTGGATTAAGATCGGGAATATCAGATTAGAG 1140
Qy 1141 GAAAGATTTTGTATGCTGAGTCCGCTGTTATACACGGGACCTTTTCACTTCTCGTT 1200
Db 1141 GAAAGATTTTGTATGCTGAGTCCGCTGTTATACACGGGACCTTTTCACTTCTCGTT 1200
Qy 1201 GCGTTTGACGGGCA 1215
Db 1201 GCGTTTGACGGGCA 1215

RESULT 2

US-09-974-300-2181
; Sequence 2181, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2181
; LENGTH: 1208
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2181

Query Match 50.8%; Score 617.2; DB 9; Length 1208;
Best Local Similarity 70.5%; Pred. No. 1.4e-185;
Matches 853; Conservative 0; Mismatches 353; Indels 4; Gaps 2;

Qy 1 ATGAGCCCAATCGATTAAATTTGTTAGTGTGCTTTCTGATCAATTTCAAAACAATCCATAT 60
Db 1 ATGATCAATCGCTAAACAATTCAGCGTGTGTGAGAACATATACAGAAATCCGTAT 60
Qy 61 GCTTATTTTCAAACTCGGGAGGAAGATCCGGTTCAATTTATGAGAGTCGATAGACAT 120
Db 61 CAGTATTTTCTATACCTTTCCGGAATCTGATCCCGTCCATTTATGAAGAATCGCTGGAC 120
Qy 121 TATTTTATCAGCGCTATCATGATGTCGCTATATCTTTCAGCATCCGATATCTTTCAG 180
Db 121 TATTTTATCAGCGCTATCATGATGTCGCGCGCTGCTCCAGAAATCAGGACGCTTTTCA 180
Qy 181 ACGAAAT--CACTTGTGAGCGTGCAGAACCAAGTATGCGAGCGCTGTGCTGGGCCAAA 238
Db 181 ACGATTTGCTGCTGCCAAAACGGCGGAGCCCGTATGCGCGGACCTGTGCTGCCCAAA 240
Qy 239 TGCATGGAAGAAACACATCTGCAAAAGAAATTTGTAGTAGAAGACTTTTATCGGTGACG 298
Db 241 TGAAGGCAAGAGACACACGCGCAAAAGAGAAATCGTTCTGCGCGCTTTTATCGGGGAAT 300
Qy 299 CACTGGATCATCTGCTCCATTTGATTAAACAAATGCAGAAAATTTGTTAGCGCCTTATC 358
Db 301 CCGTTGATCACCTCACCGCTCATCAAGAAAATGCCAAAGGCT--TTTGGCCCCGAC 358
Qy 359 TTGAAGAGGGAAGAGTATCTCGTCAATGATTTTGGAAAGAGCTTTGCGGTGTGCGTCA 418
Db 359 GTGGAGAGGCGGATGACCTTGTCAATGATTTGCGCAAAACATTTGCGCGTTTGGGTGA 418
Qy 419 CGATGGACATGCTCGGGCTGGATTAAGAGACCATGAAAAAATCTCTAGTGGCACAGCG 478
Db 419 CGATGGACATTTTAGTTTGGACAAAGACGACCAACCAACCGGTGAGAAAATGCGCACAG 478
Qy 479 GAGTTGCCGATTTTATCAGGATATCTCTCAATCTCTGAAGCGGGGACATTCGTTAT 538
Db 479 GCGTCCCGGATTTTCAATACCAGTTTGAATCAGCGCGCTGAGGACCGGGAGCAATTCCTCA 538
Qy 539 GGTGACGCAACAGCTTTCCCAATCTTCCGATGCCGCTCAATTAAGAAACCTGCGGTCAATC 598
Db 539 AATGAGTGAACAGCTCGCTGAGTATTTGAATCCGATATCGAGGAAGGCGCAAAATC 598
Qy 599 CCGGATCAGATTTAATTTTCGATCCTATGATCTTCTGAATATGAAGCATGCGGCTCGG 658
Db 599 CCGGACATGATTTAATATCCATCTTTGCACTTCCGAATACGAGGAGTGGGATGTCG 658

659	Qy	ACAAGGATATATCTCGCACTGATTTCTTAATGTGCTGTTTAGCGCGAAACGGAAACCGGCTGATA	718
659	Db	ACCGGACATACGCGCGCTGATTTCTCAATATATTGCTTGGCGGCACAGAAACCGCGGACCA	718
719	Qy	AGACGCTGGCAGCTGATGATCTACCATTTGCTCAACAATCCTGAGCAGCATGAATGATGTTT	778
719	Db	AGACGCTCGCATTAATGATTTATCATTTTGCTTTCATCATCCGATCAGATGAATGATGTGC	778
779	Qy	TGGCTGACCGTTCGTTAGTTTCCGAGAGCCATTGCGGAGACATTCGCTTATAAACCCGCCGG	838
779	Db	TGGAAGACCGCAGCGCTTCTCCCTCAGGCAATCGCAGAGACGCTCGCTCAAGCGCCGCGG	838
839	Qy	TTGAGCTGATTTCCGCGCAGCTGTCCCAAGATACAGTGTGTCGCGGTATGGAATCAAAA	898
839	Db	TGCAGCTCATTCGCGCGCAGCTTTTCACAAGACGCGGAGATTGGCGGAGTCGAGCTAAAG	898
899	Qy	AAGATACGATTGTTTTTGTATGATCGTGGCGCTTAACCGGGACCCGTGAAGCATTTTCAAC	958
899	Db	AAGGACGACTGATTTTTCATGATAGGCGCGCAATTCGCGATCTCTGAAGCGTTTCGAGG	958
959	Qy	AGCCTGACGTGTTTAAATTTATTCATCGGGAAGATCTTGGTATCAAGAGCGCTTTTAGCGCGG	1018
959	Db	ATCCGACAGTTTCAACATTTCAACGCGAGCGACTTGGGAAGTCAAAGCGCATTCAGCGCGG	1018
1019	Qy	CCGCGCGGCAATCTCGCTTTTCGGATCCCGCATTCATTAACCTGTGTAGGAGCAGCTTTTGCCA	1078
1019	Db	CAGCCAGGCATCTCGCATTCGGTTTCAGGCGTCCACAACTGTGTAGGAGCGGGGTTTCGGA	1078
1079	Qy	AAACGAAATCGAAATGTAGCTTAATTTGCTGGATAGATCGCGAAATATCAGATTAG	1138
1079	Db	AGACCGAAATCGAGCTTGTTCGCAAAATATCGTCTTGATCAGCTGAAAAATATCCGCGTGG	1138
1139	Qy	AGGAAGATTTTGTATTGCTTCAGTCCCGGTCTGTATACACGCGGACCTGTGTTTCACTTCTCG	1198
1139	Db	AGGAGACTTCATTTACCGCGAGACAGGGCTTTACACGCGGACCGCTCTCGTCAACA	1198
1199	Qy	TTGCGTTTGA	1208
1199	Db	TCGCGTTTGA	1208

RESULT 3

```

US-10-297-465A-1/c
; Sequence 1, Application US/10297465A
; Publication No. US20040142413A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Andrew
; APPLICANT: Reinach, Fernando
; APPLICANT: Setubal, Joao
; APPLICANT: Medanis, Joao
; APPLICANT: Arruda, Paulo
; TITLE OF INVENTION: Isolated Genome of Xylella Fastidiosa and Uses Thereof
; FILE REFERENCE: FAPESP 202 US (10213376)
; CURRENT APPLICATION NUMBER: US/10/297,465A
; CURRENT FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: PCT/IB01/01618
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/209,906
; PRIOR FILING DATE: 2001-06-17
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2731748
; TYPE: DNA
; ORGANISM: Xylella fastidiosa
US-10-297-465A-1

```

Db	388922	TGTTAAATGCAAGCGGTTTACCGGGCGAGAGATGAATCAATCGCTCCCTTGGATCGGATG	388763
Qy	332	ATGCAGAAAACTTGTTCAGCGCTTATCTTGAAAGAGGGAAAAAGATCTCGTCAATGATT	391
Db	388762	CGGCGCACCAATTGATCGACAACTTTGAAACAGAAACCAATCGTTGTATCTTGTGCTGAAT	388703
Qy	392	TTTGGAAAGACGTTTTCGGTGTGCGTACGATGGAACGTCTCGGGCTGATAAAGAGACC	451
Db	388702	TTTGCTTTTCCCTTTCCCATGCAGATCATTTGCCAAATGATGGATGTGATATTGGAGATG	388643
Qy	452	ATGAAAAAATCTCTGAGTGGCACAGCGGAGTTGCGGATTTTATCACGAGTATCTCTCAAT	511
Db	388642	CCGTCACCTCGGTATAGCGGTGAGCAAAATTTGCCAAGGATTTGCACCTTCGCCAATGT	388583
Qy	512	CTCTGAAGCGCGGGGACATTCGTTATGTGTGCAGGGAAACAGCTTTCCAAATCTGTATGC	571
Db	388582	CGGCCGATGAATTAGTACACGCAAGCACTGCTACGAAGAGCTTGGCAATACTATTACGA	388523
Qy	572	CGGTCAATTAAGAAACGTCGGGTCAATCCGGGATCAGATTTAAATTTTCGATCCTATGACTT	631
Db	388522	AGTTGATTGAGCTAACGGCGCACCCATCTTGAACCTGACTTAATTTCTATGTTTTCGGTG	388463
Qy	632	CTGAATATGAAGGATCGCGCTGTTCGCAAGAGATATACTCCGACTGATTTCTTAATGTGC	691
Db	388462	CTGAGGAAGACGGCGAGAAACTAACCCATGATGAAATCGTCTCCAATGTGATTAATGTTAT	388403
Qy	692	TGTTAGCGGCNAACGGAACCGCTGATTAAGACGCTGGCACTGATGATCTACCAATTTGCTCA	751
Db	388402	TAAATCGCCGGTTACGAAACCAACATCCCAATATGATTGGCAATGCAATGATGCGTTCATC	388343
Qy	752	ACAATCTGAGCAGATGAATGATGTTTGGCTGACCGGTTCTAGTTCCGAGAGCCATTG	811
Db	388342	GCCATCCAGAGCACTCGCATCTCCTCAAGAGTATCTGTCACTGATGCCACAAGCGGAT	388283
Qy	812	CGGAGACATTGCGGTTATAAACCGCCGGTTACGTGATTCCCGGGCAGCTGTCCCAAGATA	871
Db	388282	CGGAATGCTTACGTTACGACGGGTCGGTACAGTTTCAAGTCCGATGCGCGCGGATGGATGATA	388223
Qy	872	CAGTGTTCGCGGTTATGGAATCAAAAAAGATACGATTTGTTTTTTTGTATGATCGGTGGG	931
Db	388222	TAGAGTTGAAGCGCAGTTGGTCCCTCGTGGCCACAGTAGTATTTTTGATGTTGGTCTGTG	388163
Qy	932	CTAACCGGGACCCCTGAAGCAATTTGAAACAGCCTGACGTTTAAATATTCATCG	993
Db	388162	CCAACTGTATCCGGCGCAATTTCACTCATCCCGCCAGCTGGATTAATCTCGTCCGTTT	388111

RESULT 4

US-10-297-465A-1
Sequence 1, Application US/10297465A
Publication No. US20040142413A1
GENERAL INFORMATION:
APPLICANT: Simpson, Andrew
APPLICANT: Reinach, Fernando
APPLICANT: Setubal, Joao
APPLICANT: Medianeis, Joao
APPLICANT: Arruda, Paulo
TITLE OF INVENTION: Isolated Genome of Xylella Fastidiosa and Uses Thereof
FILE REFERENCE: FAPESP 202 US (10213376)
CURRENT APPLICATION NUMBER: US/10/297,465A
CURRENT FILING DATE: 2001-06-07
PRIOR APPLICATION NUMBER: PCT/IB01/01618
PRIOR FILING DATE: 2001-06-07
PRIOR APPLICATION NUMBER: 60/209,906
PRIOR FILING DATE: 2001-06-17
NUMBER OF SEQ ID NOS: 1
SOFTWARE: PatentIn version 3.2
SEQ ID NO 1
LENGTH: 2731748
TYPE: DNA
ORGANISM: Xylella fastidiosa
US-10-297-465A-1

Query Match 7.4%; Score 89.4; DB 17; Length 2731748;
Best Local Similarity 46.4%; Pred. No. 6e-15;
Matches 291; Conservative 0; Mismatches 336; Indels 0; Gaps 0;

QY 376 GATCTCGTCAATGATTTTGAAGAAGAGCTTTGCGGTGTCGCTCAGCATGACATGCTCGGG 435
DB 363766 GATCTTGTTACTGAATTTTCCCTTCCCATGCGGATCATTTGCAAGGATGATGAT 363825

QY 436 CTGGATAAAGAGACCATGAAAAATCTCTAGTGGCACACGGAGTTGCCGATTTTATC 495
DB 363826 GTGGATATCAGCGACGCTATCAGTCTAAGTGTGGCGGTGAGCAACATTGGCCAAAGGTATTG 363885

QY 496 ACAGATATCTCAATCTCTGAAGCGCGGGCACATTCTGTATGTGTGACGCAACAGCTT 555
DB 363886 GACCTTGCGCCGATGTGCGCCGATGAATTTGGTGTGACGCAAGTGTGCTTACGAAGAATTA 363945

QY 556 TCCCAATCTTGATGCGCGTCAATTAAGAAACGTGCGGTCAATCCGGGATCAGATTTAAT 615
DB 363946 GCGCATTAATCACAAGTTGATCGAGCTGCGCGCGCGCAGCTTGGAACTGACTTGAT 364005

QY 616 TCGATCTCTATGACTTCTGAATATGAAGCATGGCGCTGTGCGAACAAGATATATCTGCA 675
DB 364006 TCCATGTTGCTGCTGCTGAGGAGGAGGCCAGAAACTGACGCATGATGAAATCGTGTCC 364065

QY 676 CTGATCTTAATGTGCTGTAGCCGCAACGGACCGCTGATAAGACGCTGCACCTGATG 735
DB 364066 AATGTGATTTTGTGTTATTTGGTGGCTACGAGACCAATCCAAATGATGGCAATGCA 364125

QY 736 ATCTACCAATTTGCTCAACAATCCTGAGCAGATGAATGATTTTGGCTGACCGTTCTGTTA 795
DB 364126 TTGATTTGCTTGATCGTCAATCGAAGCAGCTTGACAGACTCAAGATGATTTGTCGTG 364185

QY 796 GTTCCGAGACCAATTCGGGAGACATTCGCTTATAACCCCGGTTTCAGCTGATTCGCGG 855
DB 364186 ATGCCGCAAGCATATTGGAATGCTTTACGTTACGATGGATCGGTACAGTTCCAGATGCGC 364245

QY 856 CAGCTGTCCCAAGATACAGTGTGCGCGGTATGGAATCAAAAAGATACGATTTGTTTT 915
DB 364246 GCGGCGATGGATGATGTGAGCATTTGAAGTGTGATAGTACCTCGTGGCAGCATTTGTTT 364305

QY 916 TGTATGATCGGTGCGCTACCGGACCTGAAGCATTTGAAACAGCTTCAGCTGTTTAAAT 975
DB 364306 TTGATGCTTGGTGTGCTTAACGCTGATCCGGCGCAATTCAGTATCCGATCATCTGGAC 364365

QY 976 ATTCTATCGGGAAGATCTTGGTATCAAG 1002
DB 364366 ATCACAGGAAGCAAGGACGCTCCAG 364392

RESULT 5

US-10-214-446-11
; Sequence 11, Application US/10214446
; Publication No. US20030180742A1

GENERAL INFORMATION:

; APPLICANT: Weiner, David
; APPLICANT: Burk, Mark J.
; APPLICANT: Hitchman, Tim
; APPLICANT: Pujol, Catherine
; APPLICANT: Richardson, Toby
; APPLICANT: Short, Jay M.

; TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING

; FILE REFERENCE: THEM AND METHODS OF MAKING AND USING THEM

; CURRENT FILING DATE: 2002-08-05

; PRIOR APPLICATION NUMBER: US 60/309,497

; NUMBER OF SEQ ID NOS: 59

; SOFTWARE: FastSeq for Windows Version 4.0

; LENGTH: 1188

; TYPE: DNA

; ORGANISM: Unknown

; FEATURE:

; OTHER INFORMATION: Obtained from an environmental sample

US-10-214-446-11

Query Match 6.7%; Score 81.4; DB 15; Length 1188;
Best Local Similarity 52.2%; Pred. No. 1.6e-14;
Matches 181; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

QY 640 GAAGCATGGCGCTGTGCGAACAAGATATATCTGCACTGATCTACATTTGCTCAACATCT 699

DB 637 GATGCGCGCCCTCGACGATCAGGCGCTGTGCGCAATATCTTTTCGCTGCGATTACG 696

QY 700 GCAACGGAAACGGCTGATAAGACGCTGGCACTGATGATCTACATTTGCTCAACATCT 759

DB 697 GGCTCGGACACGTCGCCCTGTCTATCGCGCGGGCAATCTATATCTGCGGAGCATCCG 756

QY 760 GAGCAGATGAATGATGTTTGGCTGACCGTTTCGTTAGTTCGAGAGCCATTTCGGGAGACA 819

DB 757 GCGCAGCTGGAGCGGTGCGCTCCGACCGCGGCTGATTCGCGCGCTTCGCTGAGACC 816

QY 820 TTGCGTTATAAACCCCGCGGTTGAGTTCGCGGCGAGCTGTCCCAAGATACAGTGTGTC 879

DB 817 GTGCGCTACGATCAGCCGACCAATGTACTGCGCGCGAGCTGTGCCATTGACACCGACAAA 876

QY 880 GCGCGTATGGAATCAAAAAGATACGATTTGTTTGTATGATCGGTGCGCTAACCGG 939

DB 877 TACGCAAGCGCGATGAAAAAGGTCAAGGGTCTGTTTCATGATGCGTCCGCAACCGT 936

QY 940 GACCTGAAGCATTTGAAACAGCCTGACGCTGTTTAAATATTCATCGGGA 986

DB 937 GACCGCTGGAATTCGAACACCCCGACAGTTCAATATATACCGGA 983

RESULT 6

US-09-974-300-2190

; Sequence 2190, Application US/09974300

; Patent No. US20020146721A1

; GENERAL INFORMATION:

; APPLICANT: Berka, Randy M.

; APPLICANT: Clausen, Ib Groth

; TITLE OF INVENTION: Methods For Monitoring Multiple Gene

; FILE REFERENCE: 10085.500-US

; CURRENT APPLICATION NUMBER: US/09/974,300

; CURRENT FILING DATE: 2001-10-05

; PRIOR APPLICATION NUMBER: 09/680,598

; PRIOR FILING DATE: 2000-10-06

; PRIOR APPLICATION NUMBER: 60/279,526

; PRIOR FILING DATE: 2001-03-27

; NUMBER OF SEQ ID NOS: 8481

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2190

; LENGTH: 966

; TYPE: DNA

; ORGANISM: Bacillus licheniformis

US-09-974-300-2190

Query Match 6.0%; Score 72.4; DB 9; Length 966;
Best Local Similarity 45.3%; Pred. No. 1.1e-11;
Matches 345; Conservative 0; Mismatches 411; Indels 6; Gaps 2;

QY 228 GCTGGCCCAATGATGGAAGAAGACACTCTCCAAAAGAAAGAAATTTAGTAGAAGCTT 287

DB 17 GATGATCAGCATGATGATCCCGGAAAACACACGAGAATCAGATCGATCGTCAACAGGCTTT 76

QY 288 TATCGGTGACGCACTGGGATCATCTGCTCAATTTAAACAAAATGACAGAAAATCTTGT 347

DB 77 TACACCGCGTCATGAAGAGTGGAGCCGCGCATTCGGAATGACGAAATCAATGTCT 136

QY 348 AGCGCTTATCTTGAAGAGGAAAGTATCTCGTCAATGATTTTGGAAAGACGTTTGC 407

DB 137 CGCTGATGTCCGCGCAGGGAAGAGATCGACCTTGTACAGGATTTTTCATATCCGTTGCC 196

QY 408 GGTGTCGCTCAGATGACATGCTCGGCTGGATAAAGAGACCATGAAAAATCTCTGA 467
DB 197 CGTAATCGTCATTTGCGAATGCTCGGCTTCCTTTGGTGATAAGCATCATTTTCAAGA 256
QY 468 GTGGCACAGGGGATGCGGATTTTATCAGGAG-TATCTCTCAATCTCTGAGCGCGG 526
DB 257 ATGGTCCGACCTGTTGGTCAGCTCGGAGAGCGATCGGCCGGAAGATGTGAACAGTG 316
QY 527 CACATTCGTTATGG-----TGCAGCGAAACAGCTTTTCCCAATCTTGATGCGCGTCAATAA 581
DB 317 GAAACACATCAGGACCAAGCGGAGAGAGCTGACCGCATTTCTTGAAGAGATGATTGA 376
QY 582 AGAAGCTCGGTCATTCGGGATCAGATTAATTTTCGATCCTTANGTACTTCTGAATATGA 641
DB 377 AGAGAAAGACAAAACCTCGGCAATGACCTGATTTTCGCTTTTGATCAAAAGCGGAAACAAGA 436
QY 642 AGGATGCGGCTGTCGACAGAGATATCTCGACCTGATTTCTTAATGTCTGTAGCCGC 701
DB 437 AGGGGCAAGCTGTCCCTGTATGATTTGGTTCCATTTTGGAACTCTCTTGTATGCGGG 496
QY 702 AACGGAACCGGCTGATGAAGACGCTGGCACTGATGATCTACCAATTTGCTCAACAATCCTGA 761
DB 497 GAATGAACGACGACGAAATTTAGTTTCAATGCGGCTACAGCATTTCTCGAAACACCGG 556
QY 762 GCAGATGAATGATGTTTGGCTGACCGGTTGTTAGTTTCCGAGAGCCATTTGCGGAGACATT 821
DB 557 CGTGTATGACGAGCTCGCCAGGACATCTCGAATGATTTCCGAGGACGATCGAGGAGCCGT 616
QY 822 GGTATATAACCGCGGTTGAGTGTGATTCGCGGAGCTGTCCCAAGATACAGTGTGCG 881
DB 617 CGGTTTCCGGGCGCCAGCCGATGATTTGCGCTTCGTTAAACAGGATACCGAGATCAG 676
QY 882 CGGTATGGAATCAAAAAGATACGATTTGTTTGTATGATCGGTGCGGCTAAACCGGA 941
DB 677 AGAGTAACTTGAATAAAGAGAGAGTGTGATCGCTTTCCCTTGTCTTCCCAACCGTGA 736
QY 942 CCTGAAGCATTTGAACGCGCTGACGCTGTTTAAATATTCATCG 983
DB 737 TGAACGAAATTTGAGAGGCGCACGAAATTTGATATTCACCG 778

RESULT 7

US-10-156-761-4523
; Sequence 4523, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 4523
; LENGTH: 1215
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1215)
US-10-156-761-4523

Query Match

5.6%; Score 67.6; DB 15; Length 1215;

Best Local Similarity 43.9%; Pred. No. 4.2e-10;
Matches 338; Conservative 0; Mismatches 429; Indels 3; Gaps 1;
QY 220 GGCCTGTGCTCGCCCAATGATGAAAAGAACACTCTGCCCCAAAGAAAGAAATTTGTAGTG 279
DB 265 GACCATGGGATGCTGACCTGGAGCGCGGACCAACACCGGATCAGGCGCTGGTGTGCG 324
QY 280 AGAAGCTTATCGGTGAGCAGCTGATCATCTGTCTTCCATGATTAATAAATAAATGAGAA 339
DB 325 AAGGGGTTTCAAGCGCGGACCGCTGGAGCAGCTCAAGCCGTAGTGGCCAAAGCTCGCCGC 384
QY 340 AACTTTAGTCCCTTATCTTGAAGAGGAAAGTGAATCTCGTCAATGATTTTGAAGAG 399
DB 385 GAGCTGGTGGACCGGCTGCTCGCGGCGGCGGGGATCTGCTCGCGATGTGCGCCGAG 444
QY 400 AGCTTTGCGGCTGCGCTCACGATGACATGCTCGGGCTGGATAAAGAGACCATGAAAAA 459
DB 445 CCCTCCCGTGGCGCTCATCCCGAGATGCTGGGACATCCCGAGTCCGACCGTGCCTCG 504
QY 460 ATCTCTGAGTGACACAGCGAGTTCGCGATTTTATCAAGATATCTCTCAATCTCTCGAA 519
DB 505 CTCCGCTCCCTGTCGCGGACATCTCGGAGATGTACGAGTGAACCCGC---CGAAGGAC 561
QY 520 GCGCGGACATTCGTTATGTTGTCGAGCAAGCTTTCCCAATCTTGTATGCGGTCATT 579
DB 562 GTGGCGGCAAGCGGTCGCGGCGTTCGCGAGTCTTCCGACTACTCTGCGAGCTGATC 621
QY 580 AAAGAACGTCGCGCTCAATCCGGGATCAGATTTAAATTTTCGATCTATGTACTTCTGAATAT 639
DB 622 GCGGAGCGCGCAAGAGCGCGCGGACGATCTCATCTCGGGCTCATCCCGCCCATGAC 681
QY 640 GAAGCATGCGCTGTCGAGCAAGGATATPACTCGCATGATTTCTTAATGTGCTTTAGCC 699
DB 682 GAAGCGACCGCTCACCGAGCAGGATGATCTCCACCCTGCTGCTGCTCAACCGG 741
QY 700 GCAACGGAACCGCTGATGAAGACGCTGSCATGATGATCTACCATTTGCTCAACATCTT 759
DB 742 GGCACGAGCGCCACCGCTCAACGCCACGCTCAACGCGTGGTACGCTTTGTCGCAACCCC 801
QY 760 GAGCAGATGAATGATGTTTGTGCTGACCGTTCGTTAGTTTCCGAGAGCCATTTGCGGAGACA 819
DB 802 GACGAGTGGCGCCCTGCGCGCCGACCACTGCTGCTGCGCGCCCTGGAGGATGTTG 861
QY 820 TTGCGTTATAACCGCGGTTCCAGCTGATTCGCGGCAAGCTGTCCCAAGATACAGTGGTC 879
DB 862 ATGCGCTACGACACGCCCTCCAGCTCTTCGAGCGCTGGGTCTCTGAGCAGATCGAGATC 921
QY 880 GCGGTATGGAATCAAAAAGATAGATGTTGTTTTTGTATGATCGGTGCGGCTAACCGG 939
DB 922 GACGACACGACGCTCCCGAGGGCGGAGATCGCCATGCTTTTGGCTCCGCAACCCAC 981
QY 940 GACCTGAAGCATTTGAACAGCGCTGACGCTGTTTAAATATTCATCGGGAAGA 989
DB 982 GACCCGAGGTCTTCGGGATCCCGAGAAATCGACCTCACCCCGGAGGA 1031

RESULT 8

US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089

; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match 5.6%; Score 67.6; DB 15; Length 9025608;
Best Local Similarity 43.9%; Pred. No. 1.2e-07; Mismatches 429; Indels 3; Gaps 1;
Matches 338; Conservative 0; Mismatches 429; Indels 3; Gaps 1;
QY 220 GGCCCTGTGCTGCGCCCAATGATGAAAGAACACTCTGCCAAAAGAGAAATTGTAGTG 279
DB 5543641 GACCATGGATGCTCGACCTGGAGCCCGCGGACACACCCGGATCAGGCGCTGTGTGCG 5543700
QY 280 AGAAGCTTATCGGTGACGCACTGATCATCTGTCTCCATGATTAAACAAATGCAGAA 339
DB 5543701 AAGGCGTTTACGCGCGGACCGCTGGAGCAGCTCAAGCCGTAGCTGCCAAGCTCGCCGC 5543760
QY 340 AACTTTAGCGCTTATCTTGAAGAGGAAAGTATCTCGTCAATGATTTTGGAAAG 399
DB 5543761 GAGCTGGTGGACCGGCTGCTCGCGCGCGCGGGGATCTGCTCGCGGATGTGCCGAG 5543820
QY 400 AGCTTTGCGGTGTGCTCACGATGACATGCTCGGGCTGGATAAAGAGACCATGAAAA 459
DB 5543821 CCCCTCCCGTGGCGGTCTATCGCCGAGATGCTGGGATCCCGAGTCCGACCGTGCCCG 5543880
QY 460 ATCTCTGAGTGACACAGCGGAGTTCGCGATTTTATACAGAGTATCTCTCAATCTCTGAA 519
DB 5543881 CTCCGTCCCTGCTCGCGCGGACATCTCGGGATGTACGAGCTGAACCCGCG--CGAAGGAC 5543937
QY 520 GCGCGGCACATTCGTTATGTTGCGAGCAAGCTTTCCTCAATCTTGTATGCGGCTATT 579
DB 5543938 GTGGCGCGCAAGCGCGTGTGCGGCGTGTGGAGTCTTCGATCTACCTGCGCGAGCTGATC 5543997
QY 580 AAAGAACGTCGCGTCAATCCGGGATCAGATTTAAATTTCCATCTTATGTACTTCTGAATAT 639
DB 5543998 GCGAGCGCGCAAGAGCGCGCGGACGATCTCATCTCGGGCTCATCGCCGCCATGAC 5544057
QY 640 GAAGCATGCGCTGTGCGACAAAGGATATCTCGCACTGATTTCTTAATGTGCTGTAGCC 699
DB 5544058 GAAGCGACCGCTCACCGGAGGAGATGATCTCCACCCTGCTGCTGCTCAACGCG 5544117
QY 700 GCAACGAGACCGCTGATGAAGCGCTGCGCACTGATGATCTACCATTTGCTCAACATCTT 759
DB 5544118 GCGCAGAGGCGCACCGTCAACGCCACGCTCAACGGTGTGATCGCTTGTTCGCAACCC 5544177
QY 760 GAGCAGATCAATGATGTTTTTGTGCTGACCGTTCGTTAGTTCCGAGAGCCATTCGGGAGCA 819
DB 5544178 GACCACTGCGCGCTGCGCGCGGACCACTGCTGCTGCGCGCGCGCTGGAGGATG 5544237
QY 820 TTGCGTTATAACCGCGCTTACGTTGATTCGCGCGGACGCTGTCCCAAGATACAGTGTGTC 879
DB 5544238 ATGGCTACGACAGCGCCCTCCAGCTCTTCGAGCGCTGGTCTCTGACGAGATCGAGATC 5544297
QY 880 GCGGTATGGAATCAAAAAGATAGATGTTTTTTGATGATGATCGGTGCGGCTAACCGG 939
DB 5544298 GACGCGACACCGTCCCGAGGGCGCGGAGATGCGCATGCTTTTGGCTCCGCCAACCCAC 5544357
QY 940 GACCCCTGAAGCAATTTGAACAGCGCTGACGCTGTTTAAATATTATTCGCGGAAGA 989
DB 5544358 GACCCCGAGGTCTTCCGGATCCCGAGAACTCGACCTCACCCGCGAGGA 5544407

RESULT 9
US-09-974-300-2182

; Sequence 2182, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2182
; LENGTH: 1213
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2182
Query Match 5.3%; Score 64.8; DB 9; Length 1213;
Best Local Similarity 47.0%; Pred. No. 3.3e-09;
Matches 201; Conservative 0; Mismatches 227; Indels 0; Gaps 0;
QY 560 AATACTTCATGCGGTCTAATAAGAACGTCGCGTCAATCCGGGATCAGATTTAAATTTGCA 619
DB 589 ATTATTTTAAACAACTATCAGCAGCGCAAAAAGAGCCAAAAGACGACCTGATTTTCG 648
QY 620 TCCTATGTACTTCTGAATATGAAGCATGGCGCTCTCGGACAAAGGATATCTCGCACTGA 679
DB 649 TTTTACTCGGGGCGAAGTTGACGCGCAATCGTCACAGAGAAGAACTGCTTCATTTT 708
QY 680 TCTTAATGTGCTGTAGCCGCAACGGAACCGGCTGATAAGACGCTGGCACTGATGATCT 739
DB 709 GCATCATTTCTTTGGTCGAGCAATGAGACGACAACTTGATCGCAAAACAGCGTCC 768
QY 740 ACCATTTGCTCAACAATCCTGAGCAGATGAATGTTTTTGGCTGACCGTTCGTTAGTTC 799
DB 769 GCTATCTCAGAGAAGATAAATCACAAGAAAGCGCTGAAGACAAGATCCGTCCTGCTCC 828
QY 800 CGAGAGCCATTCGCGAGACATTTGCGTTATAAACCCCGCTTCAGCTGATTCGCGCGCAGC 859
DB 829 CTGCTTTTGTGAAGAAATGCTGCGTTATATCCGCCCGTGCAGCATCGCGCCGACGG 888
QY 860 TGTCCCAAGATACAGTGTGCGCGGTATGGAATCAAAAAGATACGATGTTTTTTGTA 919
DB 889 CGGCAGAAAGACGTTGATATCGGAGCGGTGAGGATTCGAAAGGGTTCTACAGTGTACAGT 948
QY 920 TGATCGGTGCGGCTAACCGGACCCCTGAAGCATTTGAAACAGCTGACGCTGTTTAAATATTC 979
DB 949 GGGTCGCTTCAGCAATCGTGACGAATTAAGTTTGAAGATCCTGACAGCTTCAAGCTTG 1008
QY 980 ATCGGGAA 987
DB 1009 ATCGCAA 1016
RESULT 10
US-10-282-122A-25888
; Sequence 25888, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant

APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA 034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 25888
LENGTH: 1200
TYPE: DNA
ORGANISM: Mycobacterium avium
US-10-282-122A-25888

Query Match 4.6%; Score 55.6; DB 16; Length 1200;
Best Local Similarity 45.6%; Pred. No. 2.8e-06;
Matches 196; Conservative 0; Mismatches 234; Indels 0; Gaps 0;

QY	560	ATACTTGATCGCGTCAATTAAGACGTCGGTCAATCGGGATCAGATTAAATTCGA	619
Db	575	ACTACTTCTCGAGCAGATCGAACCAACGATCCCAAGCCACCGAGGACATCATCGGGG	634
QY	620	TCTATGTACTTCTGAATATGAAGCATCGCGTGTGCGACAAGGATATATCTCGCACTGA	679
Db	635	ACCTGGTACCGGGGAGATCGACGGGAAAGCTCAGCGACGAGGCCATCTACTCGTTCC	694
QY	680	FTCTTAATGTGTGTAGCCGCAACGGACCGGCTGATGAAGACGCTGGCACTGATGATCT	739
Db	695	TGCGGCTGCTGTTGCCGCGCGGCTGGAAACCACTACCGGTCTGCGGAAACCTGTTGT	754
QY	740	ACCATTTGCTCAACATCTTGACGACATGAATGATGTTTTGGTGACCGTTGTTAGTTC	799
Db	755	ATCTGCTGTCTACCAACCGCGGACCGATTCGCCCGCTGACGCGCGATCGCGAGCTGCTGG	814
QY	800	CGAGAGCCATTGCGGAGACATTCGCTTATAACCGCGGTTTCAGCTGATTCGCGGCGACG	859
Db	815	CGCGCGGATCGAGGAGGGGTTGCGGTTTCGAGACACCGCTGACCCGTCGAGCGGTTCA	874
QY	860	TGTCCCAAGATACAGTGGTTCGCGGATGGAATCAAAAAAGATACGATGTTTTTTGTA	919
Db	875	CCACCGAAGACACCGAATTCGACGGGTTCCGGATTCGCGCGGTTTCGGTGATCGCGGTG	934
QY	920	TGATCGTTCGGGTAAACCGGACCGCTGAAGCATTTGAACAGCGCTGACGTTTAAATTC	979
Db	935	GCATCGGTTTCGGCAACCGCGACGCGGCTGGGAGCGTTCCGAGGAATTCGACATCT	994
QY	980	ATCGGGAACA 989	
Db	995	TCCGTAACA 1004	

RESULT 11
US-10-229-148B-1/c
Sequence 1, Application US/10229148B
Publication No. US20040091975A1
GENERAL INFORMATION:
APPLICANT: Meiji Seika Kaisha, Ltd.
TITLE OF INVENTION: Midecamycin biosynthetic genes
FILE REFERENCE: 138451 US
CURRENT APPLICATION NUMBER: US/10/229,148B
CURRENT FILING DATE: 2002-08-28
PRIOR APPLICATION NUMBER: 210516/2002
PRIOR FILING DATE: 2002-07-19
NUMBER OF SEQ ID NOS: 48
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 84428
TYPE: DNA
ORGANISM: Streptomyces mycarofaciens
FEATURE:
NAME/KEY: CDS
LOCATION: Complement ((1)..(675))
OTHER INFORMATION: ORF42 (fragment)
FEATURE:
NAME/KEY: CDS
LOCATION: Complement ((1168)..(2202))
OTHER INFORMATION: ORF41
FEATURE:
NAME/KEY: CDS
LOCATION: Complement ((2220)..(3215))
OTHER INFORMATION: ORF40
FEATURE:
NAME/KEY: CDS
LOCATION: Complement ((3237)..(4691))
OTHER INFORMATION: ORF39
FEATURE:
NAME/KEY: CDS
LOCATION: Complement ((4695)..(5948))
OTHER INFORMATION: ORF38
FEATURE:
NAME/KEY: CDS
LOCATION: Complement ((6048)..(6629))
OTHER INFORMATION: ORF37
FEATURE:
NAME/KEY: CDS
LOCATION: Complement ((6653)..(7945))
OTHER INFORMATION: ORF36
FEATURE:
NAME/KEY: CDS
LOCATION: (8149)..(9015)
OTHER INFORMATION: ORF35
FEATURE:
NAME/KEY: CDS
LOCATION: (9012)..(9335)
OTHER INFORMATION: ORF34
FEATURE:
NAME/KEY: CDS
LOCATION: (9328)..(10458)
OTHER INFORMATION: ORF33
FEATURE:
NAME/KEY: CDS
LOCATION: (10521)..(11603)
OTHER INFORMATION: ORF32
FEATURE:
NAME/KEY: CDS
LOCATION: (11729)..(12961)
OTHER INFORMATION: ORF31
FEATURE:
NAME/KEY: CDS
LOCATION: Complement ((13016)..(14044))
OTHER INFORMATION: ORF30
FEATURE:
NAME/KEY: CDS
LOCATION: Complement ((14074)..(15096))

OTHER INFORMATION: ORP29
FEATURE:
NAME/KEY: CDS
LOCATION: (15643)..(17466)
OTHER INFORMATION: ORP28
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((17522)..(18895))
OTHER INFORMATION: ORP27
FEATURE:
NAME/KEY: CDS
LOCATION: (19063)..(20229)
OTHER INFORMATION: ORP26
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((20307)..(21743))
OTHER INFORMATION: ORP25
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((21733)..(22527))
OTHER INFORMATION: ORP24
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((22534)..(23571))
OTHER INFORMATION: ORP23
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((23555)..(24463))
OTHER INFORMATION: ORP22
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((24460)..(25650))
OTHER INFORMATION: ORP21
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((25647)..(26105))
OTHER INFORMATION: ORP20
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((26180)..(27391))
OTHER INFORMATION: ORP19
FEATURE:
NAME/KEY: CDS
LOCATION: (27937)..(28983)
OTHER INFORMATION: ORP18
FEATURE:
NAME/KEY: CDS
LOCATION: (29244)..(42779)
OTHER INFORMATION: ORP1
FEATURE:
NAME/KEY: CDS
LOCATION: (42823)..(48657)
OTHER INFORMATION: ORP2
FEATURE:
NAME/KEY: CDS
LOCATION: (48712)..(59802)
OTHER INFORMATION: ORP3
FEATURE:
NAME/KEY: CDS
LOCATION: (59850)..(64556)
OTHER INFORMATION: ORP4
FEATURE:
NAME/KEY: CDS
LOCATION: (64687)..(70365)
OTHER INFORMATION: ORP5
FEATURE:
NAME/KEY: CDS
LOCATION: (70365)..(71078)
OTHER INFORMATION: ORP6
FEATURE:
NAME/KEY: CDS
LOCATION: (71113)..(72360)
OTHER INFORMATION: ORP7

FEATURE:
NAME/KEY: CDS
LOCATION: (72400)..(73665)
OTHER INFORMATION: ORP8
FEATURE:
NAME/KEY: CDS
LOCATION: (73694)..(75043)
OTHER INFORMATION: ORP9
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((75899)..(76570))
OTHER INFORMATION: ORP10
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((76602)..(77765))
OTHER INFORMATION: ORP11
FEATURE:
NAME/KEY: CDS
LOCATION: (78039)..(79313)
OTHER INFORMATION: ORP12
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((79391)..(81052))
OTHER INFORMATION: ORP13
FEATURE:
NAME/KEY: CDS
LOCATION: (81541)..(82356)
OTHER INFORMATION: ORP14
FEATURE:
NAME/KEY: CDS
LOCATION: (82760)..(83362)
OTHER INFORMATION: ORP15
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((83495)..(84142))
OTHER INFORMATION: ORP16
FEATURE:
NAME/KEY: CDS
LOCATION: (84329)..(84428)
OTHER INFORMATION: ORP17 (fragment)
US-10-229-148B-1
Query Match 4.6%; Score 55.6; DB 16; Length 84428;
Best Local Similarity 45.4%; Pred. No. 4.1e-05;
Matches 199; Conservative 0; Mismatches 239; Indels 0; Gaps 0;
QY 546 CGAACAGCTTTCCCAATACCTTGATCCGGTCAATTAAAGAACGTCGCGTCAATCCGGGATC 605
DB 26822 CAAGGAGATGACCGACTATCTCTGCACCATCCAGATCCACGCCAAGCGCGAGGCCCGGTGCA 26763
QY 606 AGATTTAATTTTCGATCCTTGTACTTCTGAATATGAAGCATGGCGCTGTCGACAAAGCA 665
DB 26762 CGACCTGATCAGCGGCCCTGGTCCAGGCCGAGCAGACCGCCGCAAGCTCACCAGCGTGA 26703
QY 666 TATACTCGCACTGATTCTTAATGTGCTGTAGCCGCAACCGGAACCGGCTGTATAAGACGCT 725
DB 26702 GATCGTCAACATCGTGGCCCTGCTCTCACCGCCGCCACGCTCTGTCAGACGCTGCT 26643
QY 726 GGCACGTGATGATCTACCAATTTGCTCAACAATCCTGAGCAGATGAATGATGTTTGGCTGA 785
DB 26642 CAGCAACCTCTTCTCTGGTACTGGAGGAGAAACCGCAGGGCGTGGCGGACCTGCGGCCGA 26583
QY 786 CCGTTTCGTTAGTTCGAGAGCCATTGCGGAGACATTTGCTTATAACCCGCGGTTTCAGCT 845
DB 26582 CCGCGAGCTGGTGTACCGGTGGGTGGAGGAGACACTGGCGCTACCGCAGCCCTTCAACAA 26523
QY 846 GATTCCGGCGCAGCTGTCCCAAGATACAGTGTGTCGGCGCATTTGCTTATAAGATAAGATAC 905
DB 26522 CATCTTCGCTTCCTCAAGGAGGACACCGACATCTCTCGGCCCGGAATGAAGAGGCCCA 26463
QY 906 GATTGTTTTTTGATGATCGGTGCGGCTAAACCGGACCCCTGAAGCATTTTGAACAGCTGA 965
DB 26462 GATGGTTCATCGCTTGGAGCCAGTCGGCCAAACCGCGCAACCTTCCCGAGGCCGA 26403

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2004, 12:45:13 ; Search time 135.823 Seconds
(without alignments)
1069.666 Million cell updates/sec

Title: US-10-627-124-2
Perfect score: 2080
Sequence: 1 MSQIKLPSVLSDQFNPNY.....AESGLYTRGPVSLVAFDGA 405

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 359729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_23Sep04:.*
1: Geneseqp1980s:.*
2: Geneseqp1990s:.*
3: Geneseqp2000s:.*
4: Geneseqp2001s:.*
5: Geneseqp2002s:.*
6: Geneseqp2003as:.*
7: Geneseqp2003bs:.*
8: Geneseqp2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2080	100.0	405	8	ADJ62039 Bacillus
2	958	46.1	408	6	ABM70363 Phototroph
3	454	21.8	410	6	AAE36997 Micromono
4	440.5	21.2	399	6	ABP57702 Saccharop
5	438	21.1	403	8	ADM80057 Spiramyci
6	438	21.1	403	8	ADN97573 S ambofac
7	433.5	20.8	404	8	ADH39706 Streptomy
8	432	20.8	403	8	ADM45932 S. mycaro
9	430.5	20.7	404	6	ABR82126 Bacterial
10	426.5	20.5	396	3	AAE15502 Bacillus
11	423.5	20.4	404	6	ABR82143 Bacterial
12	418.5	20.1	396	3	AAE15501 Bacillus
13	417.5	20.1	396	3	AAE15503 Bacillus
14	415.5	20.0	396	2	AAE60777 Mycinamic
15	413	19.9	395	6	ABR82124 Bacterial
16	412.5	19.8	408	6	ABR82120 Bacterial
17	406	19.5	392	6	ABR82134 Bacterial
18	399.5	19.2	425	6	ABR82128 Bacterial
19	394.5	19.0	406	8	ADH39695 Streptomy
20	393	18.9	397	7	ADJ53176 Rebecamy
21	392.5	18.9	400	6	ABU34148 Protein e
22	391	18.8	418	6	AAE35502 Streptomy
23	383.5	18.4	402	6	ABU11981 M. echino
24	374	18.0	430	6	AAE34780 Streptomy
25	373.5	18.0	396	5	ABR81315 Streptomy

26	373.5	18.0	430	6	AAE34736 Streptomy
27	371.5	17.9	430	6	AAE34735 Streptomy
28	371.5	17.9	430	6	AAE34723 Streptomy
29	366.5	17.6	430	6	AAE34727 Streptomy
30	366.5	17.6	430	6	AAE34725 Streptomy
31	365.5	17.6	430	6	AAE34731 Streptomy
32	365.5	17.6	404	6	ABR82138 Bacterial
33	365	17.5	429	6	AAE34728 Streptomy
34	364.5	17.5	403	7	ADD15221 Eryf prot
35	364.5	17.5	403	8	ADO47230 S. erythra
36	362.5	17.4	424	8	ADM45925 S. mycaro
37	361.5	17.4	424	6	ABR82146 Bacterial
38	361	17.4	475	6	AAE34729 Streptomy
39	360	17.3	416	3	AAE18654 Amino aci
40	360	17.3	416	3	AAE67218 Macrolide
41	360	17.3	416	6	ABG71678 S. venezu
42	360	17.3	416	6	ADA09417 S. venezu
43	360	17.3	416	7	ADH53461 S. venezu
44	360	17.3	416	8	ADL91924 Streptomy
45	356.5	17.1	418	8	ADL39276 S. hygro

ALIGNMENTS

RESULT 1
ADJ62039
ID ADJ62039 standard; protein; 405 AA.

XX AC ADJ62039;
XX DT 06-MAY-2004 (first entry)
XX DE Bacillus subtilis cypX protein SEQ ID NO:2.
XX KW heterologous biological substance; Bacillus; cypX; red pigment.
XX OS Bacillus subtilis.
XX PN WO2004011609-A2.
XX PD 05-FEB-2004.
XX PF 25-JUL-2003; 2003WO-US023398.
XX PR 26-JUL-2002; 2002US-0398853P.
XX PA (NOVO) NOVOZYMES BIOTECH INC.
XX PI Tang M, Sloma A, Sternberg D, Behr R;
XX DR WPI: 2004-143839/14.
XX DR N-PSDB; ADJ62038.

Producing a heterologous biological substance comprises cultivating the pigment-deficient mutants of Bacillus cell that directs synthesis of the heterologous biological substance and has a modification of the cypX and ymc genes.

Example 1; SEQ ID NO 2; 62pp; English.

The present invention describes a method for producing a heterologous biological substance comprising cultivating a mutant of a parent Bacillus cell in a medium suitable for the production of a heterologous biological substance, and recovering the heterologous biological substance from the cultivation medium, where the mutant cell comprises a first nucleic acid sequence directing synthesis of the heterologous biological substance and a second nucleic acid sequence comprising a modification of at least one of the genes cypX and ymc, which are involved in the production of a red pigment. The mutant cell is also deficient in the production of the red pigment compared to the parent Bacillus cell when cultivated under the same conditions. Also described: (i) a mutant of a parent Bacillus cell, comprising a first nucleic acid sequence directing synthesis of a

Adj 10/12/03

CC heterologous biological substance and a second nucleic acid sequence
CC comprising a modification of at least one of the genes cypX and ymc,
CC which are involved in the production of a red pigment, wherein the mutant
CC cell is deficient in the production of the red pigment compared to the
CC parent Bacillus cell when cultivated under the same conditions; and (2)
CC obtaining a mutant of a parent Bacillus cell, comprising introducing into
CC the parent Bacillus cell a first nucleic acid sequence directing
CC synthesis of a heterologous biological substance and a second nucleic
CC acid sequence comprising a modification of at least one of the genes cypX
CC and ymc, which are involved in the production of a red pigment, and
CC identifying the mutant cell comprising the modified nucleic acid
CC sequence, where the mutant cell is deficient in the production of the red
CC pigment compared to the parent Bacillus cell when cultivated under the
CC same conditions. The methods and compositions of the present invention
CC are useful for producing biological substances, e.g. nucleic acid,
CC polyamide, polyamine, polyol, polypeptide, or polysaccharide, in pigment-
CC deficient mutants of Bacillus cells, and obtaining the pigment-deficient
CC mutant cells. The present sequence represents Bacillus subtilis cypX,
CC which is used in the exemplification of the present invention.

XX SQ Sequence 405 AA;

Query Match 100.0%; Score 2080; DB 8; Length 405;
Best Local Similarity 100.0%; Pred. No. 1e-196;
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSQSILFSLSDQFQNNPYAFYFQSLREDDPVHYEESIDSYFISYHVDVRYILOHPDIFT 60
DB 1 MSQSILFSLSDQFQNNPYAFYFQSLREDDPVHYEESIDSYFISYHVDVRYILOHPDIFT 60
QY 61 TKSLVERAEPVNRGVPVLAQMHGKHSKRRIIVRSFIFGDALDHLSPLIKQNAENLLAPYL 120
DB 61 TKSLVERAEPVNRGVPVLAQMHGKHSKRRIIVRSFIFGDALDHLSPLIKQNAENLLAPYL 120
QY 121 ERKSDLVNDFKTFVAVCVTMDMLGDKRDKHKISEWHSGVADFTISQSPEARAHSLW 180
DB 121 ERKSDLVNDFKTFVAVCVTMDMLGDKRDKHKISEWHSGVADFTISQSPEARAHSLW 180
QY 181 CSEQLSOYLMPIVKERRVNPQSDLSILCTSEYEGMALSDKDILALINVLAAATEPADK 240
DB 181 CSEQLSOYLMPIVKERRVNPQSDLSILCTSEYEGMALSDKDILALINVLAAATEPADK 240
QY 241 TLALMIYHLNPNQNDVLADRSIVPRAIAETLRYKPPVQLIPQLSDQTVVGGMEIKK 300
DB 241 TLALMIYHLNPNQNDVLADRSIVPRAIAETLRYKPPVQLIPQLSDQTVVGGMEIKK 300
QY 301 DTIVFCMIGAA NRDPFAEPQDVFNHREDLGKISAFSGAARHLAFGSGIHNCVGAAPAK 360
DB 301 DTIVFCMIGAA NRDPFAEPQDVFNHREDLGKISAFSGAARHLAFGSGIHNCVGAAPAK 360
QY 361 NEIEIVANIVLDKMRNIRLEEDFCYAESGLYTRGPVSLVAFDGA 405
DB 361 NEIEIVANIVLDKMRNIRLEEDFCYAESGLYTRGPVSLVAFDGA 405

RESULT 2

ABW70363

ID ABW70363 standard; protein; 408 AA.

XX AC

XX ABW70363;

XX 20-NOV-2003 (first entry)

DE Photorhabdus luminescens protein sequence #3460.

XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW detection; food; gene expression; plant; animal; microorganism; toxin;
KW antibiotic; biopesticide; virulence factor; disease model; plague;
KW whooping cough.

XX Photorhabdus luminescens.

OS WO200294867-A2.

PN

XX 28-NOV-2002.
PD 07-FEB-2002; 2002WO-IB003040.
XX 07-FEB-2001; 2001FR-00001659.
PR (INSP) INST PASTEUR.
XX (CNRS) CNRS CENT NAT RECH SCI.
PA Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
PI Buchrieser C;
XX WPI; 2003-148459/14.
XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
PT Claim 2; SEQ ID NO 3460; 1205pp; French.
XX The invention relates to the isolation of genes and their encoded
CC proteins from Photorhabdus luminescens. The isolated sequences are
CC sources of probes and primers for detecting the genome of P. luminescens
CC and related species; to study polymorphisms; for gene analysis and for
CC detection/amplification of the genes. Antibodies (Ab) raised against the
CC polypeptides encoded by the genes are used for detection/identification
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC carry a gene-containing vector are used to select compounds that
CC modulate, regulate, induce or inhibit expression of the genes in plants,
CC animals or microorganisms other than P. luminescens and are able to alter
CC response or sensitivity to toxins and antibiotics produced by P.
CC luminescens. Cells transformed to express the genes are useful for
CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luminescens proteins
XX SQ Sequence 408 AA;

Query Match 46.1%; Score 958; DB 6; Length 408;

Best Local Similarity 46.6%; Pred. No. 1.3e-85;

Matches 185; Conservative 82; Mismatches 124; Indels 6; Gaps 3;

QY 8 FSVLSDDQFQNNPYAFYFQSLREDDPVHYEESIDSYFISYHVDVRYILOHPDIFTTKSLVER 67

DB 7 FNHDPKFKNPYDFYDLHKQDLVYFQSQNSYFQYGVKEDVDALK-SSIFNTKPLTAL 65

QY 68 AEPVNRGVPVLAQMHGKHSKRRIIVRSFIFGDALD-HLSPLIKQNAENLLAPYLRGKSD 126

DB 66 AEPVNGDRVLAQMEGEEHACKRKFIMQGLSRDYFNRYEYPMIRKITEDLLQPYMEKGNID 125

QY 127 LVNDPQKTFVAVCVTMDMLGDKRDKHKISEWHSGVADFTISQSPEARAHSLSEQLS 186

DB 126 LVNDPGRDYAVLVTLSILGLPSDNYRDIAEWHKGIASFITQDQTELEKRMHSLSCSKIL 185

QY 187 QYLMPIVKERRVNPQSDLSILCTSEYEGMALSDKDILALINVLAAATEPADKTLALMI 246

DB 186 RLLKPIIDORRNPQSDLSIFC----QDTAMNSEITALCLNILLAAATEPADKILAMWL 241

QY 247 YHLLNPNQNDVLADRSIVPRAIAETLRYKPPVQLIPQLSDQTVVGGMEIKKDTIVFC 306

DB 242 NLLISNPSMLDVLKDRSLVADAEETLRLTSPVQLIPREASEDVTISGIDIPKGAUVFC 301

QY 307 MIGANRDPFAEPQDVFNHREDLGKISAFSGAARHLAFGSGIHNCVGAAPAKNEIEIV 366

DB 302 MIGANRDPVFFHKPNFEFDLYRRKNTTSPQKARRKRLHAFAGTHACAAAFAFSLSEQLS 361

QY 367 ANIVLDKMRNIRLEEDFCYAESGLYTRGPVSLVAFD 403

Db 64 AAAILNRDVRGFPFLILREHSLGTMPPPEHTRLRKLVGKAFARRVEQLRPRQQLVDHL 123
QY 116 LAPYLERG-KSDLVNDFGKTFACVCTMDMLGDKDHEKISWHSGVADFTISQSPEA 174
Db 124 LORMAADGPGDLVSALALPLPIKVICDLGIPVADRFRVW-SDIALAIT--SNSPEE 180
QY 175 RAHSLWCSQLSQYLMPVIKERRVNPGLISILCTSEYEGMALSDKOILALINVLAA 234
Db 181 IRES---RQIYAYIGELVQQRKKPTEDLLSVLQVAREAGQLSEELVTCAGLLIAG 237
QY 235 TEPADKTLALMIYHLNPNQMDVADRSLVPRAIAETLYKP--PVQLIPRQLSQDVT 292
Db 238 FETTANHIANFTFNLLTHPDQLKLIADPELVPAVEELLRYTLGATPGFPRIATEDLE 297
QY 293 VCGMEIKKDTIVFCMIGANRDPFAEQPDVFNHREDLGKISAFSGAARHLAFSGHIN 352
Db 298 LGGVSTRGDVAFVFEIASANRDSAVFDGPDLDLAREH-----NSHMALGHGPHY 347
QY 353 CVGAFAKNEIEIVANIVLDKRNII--RLEEDFCVAESGLYTRGVPVSLLVAF 402
Db 348 CIGQLARMELOVAGTLIKRFPQLSFVFPVDEVVWVWKRGMTRGPEALPITW 399
RESULT 5
ADM80057
ID ADM80057 standard; protein; 403 AA.
AC ADM80057;
XX
XX
DT 01-JUL-2004 (first entry)
DE Spiromycin biosynthesis orf1 protein, SEQ ID 24.
XX
KW Antibacterial; Immunosuppressive; Antirheumatic; Antiarthritic;
KW Anthelmintic; Insecticide; spiramycin biosynthesis; macrolide;
KW spiramycin; antibiotic; immunosuppressant; autoimmune disease.
XX
OS Streptomycetes ambofaciens.
XX
XX FR2845394-A1.
XX
XX 09-APR-2004.
XX
XX 08-OCT-2002; 2002FR-00012489.
XX
XX 08-OCT-2002; 2002FR-00012489.
XX
XX (AVET) AVENTIS PHARMA SA.
XX (CNRS) CNRS CENT NAT RECH SCI.
XX
XX Blondelet RMH, Dominguez H, Darbon RE, Gerbaud C, Gondran A;
XX Karray F, Lacroix P, Oestreicher MBN, Pernodet JL, Tuphile K;
XX
XX WPI: 2004-307291/29.
XX N-PSDB; ADM80056.
XX
XX New polynucleotides encoding proteins involved in spiramycin
XX biosynthesis, useful for improving synthesis of macrolide antibiotics or
XX for generating new hybrid macrolides.
XX
XX Claim 8; SEQ ID NO 24; 323pp; French.
XX
XX The present invention relates to coding sequences (I) which encodes
XX proteins involved in spiramycin (A) biosynthesis. (I) are used to improve
XX production (optionally also purity) of macrolides (M), especially
XX spiramycins, and to produce hybrid antibiotics. (M) are useful in human
XX and veterinary medicine as antibiotics, but some are also
XX immunosuppressants (e.g. FK506) useful for treating transplant rejection,
XX rheumatoid arthritis and other autoimmune diseases; or insecticides and
XX anthelmintics (e.g. avermectin). The present sequence is one such
XX protein involved in spiramycin biosynthesis.
XX

SQ Sequence 403 AA;
Query Match 21.1%; Score 438; DB 8; Length 403;
Best Local Similarity 29.2%; Pred. No. 4e-34;
Matches 110; Conservative 77; Mismatches 156; Indels 34; Gaps 7;
QY 22 YFSQREDDPVHYBESIDSFYISRYHVRVYILOHPDIFTTK-SLVERAEPVNR---GP-V 76
Db 23 WFAYNRTRHPVFWDETRQAWQVFGYDDVYTVNNPQPFSSDFNMVMPPTPELEMIIGPT 82
QY 77 LAQMHGKHSKRRIIVRSFGDLDHLSPLIKQNAENLLAPYLERGKSDLVNDFGKTPA 136
Db 83 ICAIDPPAHGPMKLVLSQAFTFRRIARLEPRVRAITTELLDKVGQDDVVDVAGDLSYALP 142
QY 137 VCVTMDMLGLOKROHEKISEW-----HSGVADFTISQSPEARAHSLWCSEOL 185
Db 143 VIVIAELGIGIPAGRDLPREWVDTLTNEGLEYNLPDNFT-----ETIAPAL---KEM 193
QY 186 SQYLMFVIKERRVNPGLISILCTSEYEGMALSDKOILALINVLAAATEPADKTLALM 245
Db 194 TDYLLKQIHAKRDAPADLVSLVQAEQDGRRLTDVEIVNIVALLLTAGHVSSTLLSNL 253
QY 246 IYHLNPNQMDVADRSLVPRAIAETLYKPVPQLIPRQLSQDVTVCGMEIKKDTIVF 305
Db 254 FLVLEENPQALEDLRADRSLVFGAIEETLYRSPFNFIIRFYKEDTTVLGPLMERGQWVI 313
QY 306 CMIGAANDPFAEQPDVFNHREDLGKISAFSGAARHLAFSGHINCVGAFAKNEIEI 365
Db 314 AWSQANRDRPHFPDPDPTDFIRSD-----GTRHWAFGHGHICLGAALARLEGV 364
QY 366 VANIVLDKVRNIRLEED 382
Db 365 MLELLDRVQGFRI DHE 381
RESULT 6
ADM97573
ID ADM97573 standard; protein; 403 AA.
AC ADM97573;
XX
XX 15-JUL-2004 (first entry)
DE
XX
XX S ambofaciens spiramycin biosynthetic enzyme encoded by ORF1.
KW antibacterial; immunosuppressive; antirheumatic; antiarthritic;
KW anthelmintic; insecticide; biosynthesis; macrolide; antibiotic;
KW spiramycin; veterinary medicine; immunosuppressant; transplant rejection;
KW rheumatoid arthritis; autoimmune disease; insecticide; anthelmintic;
KW avermectin.
XX
XX Streptomycetes ambofaciens.
XX
XX WO2004033689-A2.
XX
XX 22-APR-2004.
XX
XX 08-OCT-2003; 2003WO-FR002962.
XX
XX 08-OCT-2002; 2002FR-00012489.
XX 27-FEB-2003; 2003FR-00002439.
XX 07-AUG-2003; 2003US-0493490P.
XX
XX (AVET) AVENTIS PHARMA SA.
XX (CNRS) CNRS.
XX
XX Blondelet-Rouault M, Dominguez H, Darbon-Rongere E, Gerbaud C;
XX Gondran A, Karray F, Lacroix P, Oestreicher-Mermet- Bouvier N;
XX Pernodet J, Tuphile K;
XX WPI: 2004-330455/30.
XX DR N-PSDB; ADM97572.
XX


```
XX 03-JUN-2004 (first entry)
DT ABR82126
XX ABR82126 standard; protein; 404 AA.
DE AC ABR82126;
XX 30-SEP-2003 (first entry)
KW DT Bacterial P450 enzyme SEQ ID NO:16.
XX DE Bacterial; P450; enzyme; catalysis; hydrolysis; epoxide; arene oxide;
XX KW diol; alkene; chirality; thermotolerance; thermostability.
XX OS Eubacteria.
XX PN WO2003052050-A2.
XX PD 26-JUN-2003.
XX PF 05-AUG-2002; 2002WO-US024910.
XX PR 03-AUG-2001; 2001US-0309497P.
XX PA (DIVE-) DIVERSA CORP.
XX PI Weiner D, Burke M, Hitchman T, Pujol C, Richardson T, Short J;
XX WPI; 2003-541641/51.
XX DR N-PSDB; ACF06116.
XX PT Novel polypeptide having P450 enzyme activity and polynucleotides
XX encoding the polypeptide, useful for catalyzing the hydrolysis of
XX epoxides and arene oxides to their corresponding diols.
XX PS Claim 104; Page 58-59; 365pp; English.
XX CC The present sequence represents a bacterial P450 enzyme. P450 enzymes can
XX be used to catalyse the hydrolysis of epoxides and arene oxides to their
XX corresponding diols. P450 enzymes can also be used for hydrolysing an
XX alkene, for producing a compound of a desired chirality, and for
XX increasing thermotolerance or thermostability of a P450 polypeptide
XX Sequence 404 AA;
XX Query Match 20.7%; Score 430.5; DB 6; Length 404;
XX Best Local Similarity 28.8%; Pred. No. 2.2e-33;
XX Matches 114; Conservative 81; Mismatches 164; Indels 37; Gaps 11;
QY 24 SOLREEDPV-----HYESIDSYFISRYHDVRYILQHPDIFTTKSLVERABPMRGPVLA 78
DB 29 TELLEKEPVAIRLPFAARNEAWLVRYEDVRAVTS DPR-FSRTALLQQVTKMGHVA 87
QY 79 QMHGKE-----HSAKRIVVRSFIGDALDHLSPLIKQNAENLLAPYLERGK-SDLVND 131
DB 88 SKAALNVADPPYHTQLRKAVTKAFTGQSTRRLRPLAQAGTDRLLDAMEAAGRPAADMKHL 147
QY 132 GKTFAVCVTMDMLGDKDEHEKISEHSGVADFTISISQSPEARAHLWCSEQLSOYLMP 191
DB 148 HGPLPMVAVCDLLGIPEDRAELASW-----POLILSSGGPES-----SKAAQAIHGYVIR 200
QY 192 VIKERRVNPGLDLSILCTSEYEGMALSDKDLILNLVLAATEPADKTLALMIYHLIN 251
DB 201 LLDRRAEPQDLAGVLAESLAEG-RITAEAVSLAMAILISGAHAVRNSANWVYVLLT 259
QY 252 NPEQMDVLADRLSVRAIAETLYKPPVQLI--PRQLSQDTVVGGMEIKKDTIVFCMIG 309
DB 260 RPELADRLRAEPGLLPQAVDELLRWPHRNGVGLPRIATEDVEVGGVILIRAGEAVVASYL 319
QY 310 AANRDPFAEPQDVFNTHREDLGIKSAFGAARHLAFSGIHNCVGAFAAKNEIEIVANI 369
DB 320 AANRDPFAAFEDFLDFDREGIG-----HVSFGHPHHCMAMLTRMSEVMUST 369
QY 370 VLDKMRNIRL---BEDFCYAESGLYTRGFSLLVAF 402
```

```
XX 03-JUN-2004 (first entry)
DT S. mycarofaciens midcamycin polyketide synthetase ORF19 protein.
DE KW macrolide; midcamycin biosynthesis; polyketide synthetase; PKS; enzyme;
XX antimicrobial; ORF19.
XX OS Streptomyces mycarofaciens.
XX PN JP2004049100-A.
XX PD 19-FEB-2004.
XX PF 19-JUL-2002; 2002JP-00210516.
XX PR 19-JUL-2002; 2002JP-00210516.
XX PA (MEIJ ) MEIJI SEIKA KAISHA LTD.
XX WPI; 2004-174115/17.
XX DR N-PSDB; ADM45913.
XX PT New polynucleotides encoding proteins important in the biosynthesis of
XX midcamycin.
XX PS Claim 1; SEQ ID NO 20; 264pp; Japanese.
XX CC The invention relates to a novel polynucleotide, specifically a
XX biosynthesis gene, comprising a nucleotide sequence which encodes a
XX protein, where the gene is concerned with the biosynthesis of
XX midcamycin. The polynucleotide of the invention demonstrates
XX antimicrobial activity and may be useful for the biosynthesis of
XX midcamycin, as well as for manufacturing macrolide compounds other than
XX midcamycin. The current sequence is that of the Streptomyces
XX mycarofaciens midcamycin polyketide synthetase (PKS) ORF19 protein of
XX the invention.
XX Sequence 403 AA;
XX Query Match 20.8%; Score 432; DB 8; Length 403;
XX Best Local Similarity 30.0%; Pred. No. 1.6e-33;
XX Matches 111; Conservative 76; Mismatches 159; Indels 24; Gaps 6;
QY 22 YPSQLREEDPVHYESIDSYFISRYHDVRYILQHPDIFTTKSLVERABP-----VMRGP-V 76
DB 23 WFAFNRTHPVFWDESRHAWQVRYDDYLTGVSNNPQFFSSDFNEVMPTPELEMVIGPQT 82
QY 77 LAQMHGKHSAXRRIVVRSFIGDALDHLSPLIKQNAENLLAPYLERGKSDLVNDFGKTF 136
DB 83 IGALOPPAHGPMPKLVLSQAFTPRMARLEPRIRAVTQELLDVAVRQETIDVVGDLISYALP 142
QY 137 VCVTMDMLGLDKRDHEKISEW-----HSGVADFTISISQSPEARAHLWCSEQLSOYL 190
DB 143 VIVIAELLGIPSGDRDVRFGWVDTLTNEG-L-EYPNLPNFSSETIAPAL---KEMTDYLL 198
QY 191 PVIKERRVNPGLDLSILCTSEYEGMALSDKDLILNLVLAATEPADKTLALMIYHL 250
DB 199 HQIHAKREAPVDLLISGLVQAEQDGRKLTDEIVNIVALLTAGHVSSSTLSNLFVLV 258
QY 251 NNPEQMDVLADRLSVRAIAETLYKPPVQLIPQLSQDTVVGGMEIKKDTIVFCMIGA 310
DB 259 ENPQALADLRARELVTGAVETLYRSPFNNIFRFLKEDTDILPEMKKGQMWIAWSQS 318
QY 311 ANRDPFAEPQDVFNTHREDLGIKSAFGAARHLAFSGIHNCVGAFAAKNEIEIVANIV 370
DB 319 ANRDPHEPPEPDTDIR-----SSSRHMAFGIHHCLGAFLARQSGKVVL 369
QY 371 LDKMRNIRLE 380
DB 370 LDQVREFRID 379
```

[illegible]

Db	82	ISM	PPKHTKIRSVVNKAFTPRANKQWEPRIQETIDELIOKFOGRSEFFDLVHDYSYPLPV	14
Qy	138	CVT	MDMLGDKDRDHEKISEWHSQVADFTTISQSPPEARASLW-----CSEQLSOYLM	191
Db	142	IVISELLG	WPSAHMEQFKAW----SDLLVSTPKDSEAEKAFLEERDKCEELAAFFAG	197
Qy	192	VIKERRV	NPVSGDLISILCTSEYEGMAISDDKILAILINVLAAATEPADKTLALMIYHLIN	251
Db	198	ITIERKRN	PEQDIISILYAEAEETGKLSGEELIPLCFLLLVAGNETTTNLLISNAFMSILE	257
Qy	252	NPEQMN	VDLAQRSVPRAIAETLRYPKVPVQLIPROLSQDTWVGMEIKDITIVFCMIGAA	311
Db	258	TGCVTEEL	RSHPELMPQAVEALFRAPAPVIRRIAKRDTEIGHLIKEGDVLAFVASA	317
Qy	312	NRDPEAF	EQPDVFNTHREDLGIKSAFSGAARHLAFSGGIHNCVGAFAKNEIEIVANIYL	371
Db	318	NRDEAK	FRPHMFDIRRH-----PNPHIAFGHGHIFCLGAPLARLE---ANIAL	363
Qy	372	DKM	374	
Db	364	TSL	366	
RESULT 14				
AAR60777				
ID	AAR60777 standard; protein; 396 AA.			
AC	AAR60777;			
DT	16-OCT-2003 (revised)			
DT	21-JUN-1995 (first entry)			
XX	Mycinamicin IV hydroxylating protein.			
DE	XX			
XX	Mycinamicin; hydroxylase; macrolide; antibiotic; Micromonospora;			
KW	Micromonospora griseorubida.			
KM	XX			
XX	Micromonospora griseorubida; AT11725CN3.			
OS	JP06253853-A.			
PN	13-SEP-1994.			
PD	XX			
XX	09-MAR-1993; 93JP-00047638.			
PF	XX			
PR	09-MAR-1993; 93JP-00047638.			
XX	(ASAH) ASahi KASEI KOGYO KK.			
PA	XX			
WPI	1994-328997/41.			
DR	N-PSDB; AAQ73674.			
XX	DNA encoding a protein having mycinamicin IV hydroxylating activity - for			
PT	prodn. of mycinamicin, a macrolide antibiotic.			
XX	XX			
PS	Claim 1; Page 12-14; 23pp; Japanese.			
XX	XX			
CC	The amino acid sequence of a protein having mycinamicin IV hydroxylase			
CC	(MH) activity. The gene encodes a protein of 396 a.a. The DNA was			
CC	obtained from the macrolide antibiotic-producing bacterium Micromonospora			
CC	griseorubida AAA11725CN3/PTYS507. The gene was isolated from the plasmid			
CC	PTYS507. The protein encoded by this plasmid can be used to produce			
CC	mycinamicin IV in PTYS507-deficient Micromonospora strains. (Updated on			
CC	16-OCT-2003 to standardise OS field)			
XX	XX			
Sequence	396 AA;			
Query Match	20.0%; Score 415.5; DB 2; Length 396;			
Best Local Similarity	28.9%; Pred. No. 6.6e-32;			
Matches 114; Conservative	76; Mismatches 171; Indels 33; Gaps 11			
Qy	23	FSQLEEDPVH-----YVESIDSVFISRYHDVRYLOHQHPDFTTKSLVE---RAEPPVMRG	74	

		Matches	112;	Conservative	72;	Mismatches	162;	Indels	46;	Gaps	8;
Qy	8	FSVLSDFQFQNNPYAVFSQ	REDPVHYEES	IDSYFISRYHDVRYILQHPDIETT-----	61						
Db	5	FNPYCEEFTQNPQWQFRA	LTKQDPVHYIEEFDAWALFGFEDVWRAGMDRESFATYGGSP	64							
Qy	62	-KSLVERAEPVNRGVPVLA	QMHGKHSAKRIYVRSFIGDALDHLSP	LKQNAENLLAPYL	120						
Db	65	QALLIDR---	VKQPEIFLPMIDIPNMIH	RGIITAKDYGRNAMP	LLGKIRATAKEAITPYL	121					
Qy	121	ERKSDLVNDFGKTFPACV	TMDMLGLDKRDHEKI-----	SEWHS	GVADFTISISQ	170					
Db	122	KSGEMD-VYAFARTVALE	FTIADMIGLRPEEVVRI	RS	LIDIFFGRTFGRHGT-----	171					
Qy	171	SPEARAHSLWCSEQLSQ	YLMPIVKERRVN--	PGSDLISILCTSEYEG	WALS	DKDILALIL	228				
Db	172	TPDGVV----	AFHEVTAYVLDLIGHY	RAKGAPEGSHIDNWLK	AEPDGRPLDDQALCANIF	227					
Qy	229	NVLLAATEPADKTLALMI	YHLLNNPEQMN	DVLADRS	LVPRAJAETLR	YKPPVQLIPROLS	288				
Db	228	SLSITGSDTVPLSSAA	IYYLSEHPAQLEAV	RS	DRA	LIPAAFAETVRYDQPTNV	LGRLIA	287			
Qy	289	QDTVVGMEIKKDTIVFC	MIGANRDPEAF	QOPDVFN	IHREDLG	IKSAFSGAARHLARGS	348				
Db	288	IDTDKYGKPMKKGQAV	LFWYASANRDP	LEFEHPTF	FN	IYRD-----	PRRTLSFGS	337			
Qy	349	GIHNCVGAFAFAKNEI	EVANIVLDK	MNIRLE	380						
Db	338	GIHICIGQLLAKLEGQ	ILILETFEHI	PFTVQ	369						

Search completed: December 10, 2004, 13:40:37
Job time : 138.823 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2004, 13:27:20 ; Search time 34.5732 Seconds
(without alignments)
776.868 Million cell updates/sec

Title: US-10-627-124-2
Perfect score: 2080
Sequence: 1 MSQSLFSLVSDQFNPNP.....AESGLYTRGPVSLVAFDGA 405

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A-COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B-COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A-COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B-COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PTUS-COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	383.5	18.4	402	US-09-724-797-36	Sequence 36, Appl
2	360	17.3	416	US-09-320-878-18	Sequence 18, Appl
3	360	17.3	416	US-09-105-537-39	Sequence 39, Appl
4	360	17.3	416	US-09-141-908-13	Sequence 13, Appl
5	360	17.3	416	US-09-657-440-18	Sequence 18, Appl
6	353.5	17.0	399	US-08-765-907A-10	Sequence 10, Appl
7	352.5	16.9	419	US-09-335-409-8	Sequence 8, Appl
8	352.5	16.9	419	US-09-413-814-71	Sequence 71, Appl
9	352.5	16.9	419	US-09-568-102-8	Sequence 8, Appl
10	352.5	16.9	419	US-09-567-969-8	Sequence 8, Appl
11	352.5	16.9	419	US-09-568-480-8	Sequence 8, Appl
12	352.5	16.9	419	US-09-568-486-8	Sequence 8, Appl
13	352.5	16.9	419	US-09-568-472-8	Sequence 8, Appl
14	352.5	16.9	419	US-09-567-899-8	Sequence 8, Appl
15	347.5	16.7	406	5212296-6	Patent No. 5212296
16	335	16.1	403	5212296-9	Patent No. 5212296
17	335	16.1	448	US-09-724-797-44	Sequence 44, Appl
18	302	14.5	437	US-09-252-991A-17836	Sequence 17836, A
19	273.5	13.1	412	US-08-102-863-11	Sequence 11, Appl
20	273.5	13.1	412	PCR-US92-10885-11	Sequence 11, Appl
21	271	13.0	409	US-09-385-028-12	Sequence 12, Appl
22	271	13.0	409	US-09-726-614-12	Sequence 12, Appl
23	271	13.0	409	US-09-385-040-12	Sequence 12, Appl
24	257	12.4	468	US-09-252-991A-32437	Sequence 32437, A
25	249.5	12.0	395	US-09-266-965-129	Sequence 129, Appl
26	238	11.4	189	US-09-679-279-20	Sequence 20, Appl
27	221.5	10.6	504	US-09-126-420A-18	Sequence 18, Appl

28	213	10.2	524	4	US-09-976-594-533	Sequence 533, Appl
29	197	9.5	444	3	US-09-413-814-95	Sequence 95, Appl
30	197	9.5	444	3	US-09-413-814-96	Sequence 96, Appl
31	195	9.4	414	4	US-10-018-730A-2	Sequence 2, Appl
32	195	9.4	509	4	US-10-140-002-108	Sequence 108, Appl
33	192.5	9.3	510	4	US-09-852-067-2	Sequence 2, Appl
34	192	9.2	513	4	US-09-142-108C-15	Sequence 15, Appl
35	192	9.2	526	1	US-08-298-426-4	Sequence 4, Appl
36	188	9.0	489	4	US-09-852-067-4	Sequence 4, Appl
37	186	8.9	476	4	US-09-142-108C-23	Sequence 23, Appl
38	184.5	8.9	484	4	US-10-142-231-56	Sequence 56, Appl
39	183	8.8	496	3	US-09-172-339-6	Sequence 6, Appl
40	182.5	8.8	503	4	US-09-583-447A-2	Sequence 2, Appl
41	182	8.8	496	3	US-09-292-768-64	Sequence 64, Appl
42	182	8.8	500	3	US-09-292-768-68	Sequence 68, Appl
43	178.5	8.6	422	1	US-08-396-218-2	Sequence 2, Appl
44	178.5	8.6	422	1	US-08-760-116-2	Sequence 2, Appl
45	178	8.6	496	3	US-08-881-784-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-724-797-36
; Sequence 36, Application US/09724797
; Patent No. 6733998
; GENERAL INFORMATION:
; APPLICANT: Jon S. THORSON
; TITLE OF INVENTION: MICROMONOSPORA ECHINOSPORA GENES
; TITLE OF INVENTION: ENCODING FOR BIOSYNTHESIS OF
; FILE REFERENCE: 2653-40
; CURRENT APPLICATION NUMBER: US/09/724,797
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 60/111,325
; PRIOR FILING DATE: 1998-12-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 36
; LENGTH: 402
; TYPE: PRT
; ORGANISM: Bacteria.
US-09-724-797-36

Query Match	18.4%	Score 383.5;	DB 4;	Length 402;
Best Local Similarity	26.7%	Pred. No. 1.6e-32;		
Matches 111;	Conservative 87;	Mismatches 180;	Indels 37;	Gaps 13;
QY	4	SIKLSVLSQFNPNP	YAFSOLREEDPVHYVESIDSYPISRYHDVRYILOHPDIFTTKS	63
Db	5	AVTAFDPTDADVRDRDP	PSYHMLLRHDPVH-RGAHRVWVVSFADVRAVL--GDERFART	61
QY	64	LVERAEFVNRGP-VLAQMHG	-----KEHSAKRIVVRISFGDLDHLSPLIKQNAEN	114
Db	62	GIRREFTDLVGLAEI	VGDIILFQDEPDHGLRGVGPAPSPALRLEFVIATVD	121
QY	115	LIAPYLERKSDLVND	FGKTFACVCTMDLGDKEHKEISEWHSGVADFI-TSISQSPE	173
Db	122	LURPALAGMDDVDE	LAYPLRALVGLGLLPADWAGVGRWSRDVGTLDGSAEOM	181
QY	174	ARAHSLWCSEQLSQ	YLMPIKERRVNPGRSDILSTSEYEGMALSDKDILALINLVLA	233
Db	182	REGHA--AIAEPAD	IVERALARRRREGGDLALMLADHRLG--NSREIVSTVTFT	238
QY	234	ATEPADKTLALMI	YHLNPNPEQMNVDLADR-SLVPRATAETURYKPPVQLIPROLSQDTV	292
Db	239	GHETVASQVGN	AVLSLLAHPDQL-DLLRRRPLLAQAVEECLRYDPVSQSNTRQLDDVYE	297
QY	293	VGMEIKKDTIV	FCMIGANRDPFAEQPDVNIHREDIGIKSAFSGAARHIAFGSGIHN	352
Db	298	LRGRRLRRDDVV	VVLAGAANRDRPRYDRDPDPDIERDPV-----PSMSFGAGMY	347


```
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: PROV. 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-141-908-13

Query Match
Best Local Similarity 17.3%; Score 360; DB 4; Length 416;
Matches 102; Conservative 76; Mismatches 173; Indels 30; Gaps 8;

QY 11 LSDQFQNNPYAFYFSQREEDPVH---YBESIDSYFISRYHVDVRYLIQHP-----DIPTTKS 63
D 20 LQDPAADPYPTIYARLAEGPAHRVTPGDEVLVVGVDRAVLAADPRFSKDWNRNSTT 79
QY 64 LVERAEPVMRGVPLAQMKGKHSKRRIIVRSFISGIDALDHLSPLIKQNAENLLAPYLER- 122
D 80 PLTEAALNNHNL-ESDPPRTRLRKLVAREFTMRRELLRPRVQEIVDGLVDAMLAAP 138
QY 123 -GKSLVNDFGKTFVAVCTMDMLGDKRDHEKISEWHSGVADFITSISQSPPEARHSLWC 181
D 139 DGRADLMESLAWPLPITVISELGVPEPDRAAFRW-----TDAFVFPDDPAQAQTA 190
QY 182 SEQLSOYLMVPIKERRVNPGLSILC-TSEYEGMALSDKDILAILNLVLLAATEPADK 240
D 191 MAEMGYLSRLDSKRGQGEDLLSALVTSDEDSRLTSEELGMAHILLVAGHETTIVN 250
QY 241 TLALMIYHLNNPEQNDVLADRSVPRAIATLRYKPPVQIIPQLSQDTPV-VGMEIK 299
D 251 LIANGMYALLSHPDQALADMTLLDGAVEMLRYEGVESATYRFPVPVLDGTVIP 310
QY 300 KDTIVFCMIGANRDEPEAQPDVFNHREDIGIKSAFSGAARHLAFSGIHCNCGAFA 359
D 311 AGDTVLVLADAHRTPERFPDHRDIRDTAG-----HLAFGHIHFCIGAPLA 360
QY 360 KNEIEIVANIVLDKMRNIRLE 380
D 361 RLEARIARALLERCPLDALD 381

RESULT 6
US-08-765-907A-10
; Sequence 10, Application US/08765907A
; Patent No. 6352839
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAMAS-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: DUTRUC-ROSSET, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE OF INVENTION: Mutasynthesis
; CURRENT APPLICATION NUMBER: US/08/765,907A
; CURRENT FILING DATE: 1997-03-20
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 399
; TYPE: PRT
; ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-10

Query Match
Best Local Similarity 17.0%; Score 353.5; DB 3; Length 399;
Matches 107; Conservative 65; Mismatches 175; Indels 63; Gaps 10;

QY 21 AYFSOLREEDPVHYBESIDSYFISRYHVDVRYLIQHPDIFTTK-SLVERAEPVMRGVPLAQ 79
D 18 AMLREWRHHFVH-EDEYCAFHVFRHADVLTVASPGVYSSQLSRILPQSQALSEQILSV 76
QY 80 MHGKHSKRRIIVRSFISGIDALDHLSPLIKQNAENLLAPYLERGKSDLVNDFGKTFVAVC 139
D 77 IDPPMHRTLRRLVSQAFTPRTVADLEPRVTELQGLL-DAVDGDTFDLVADEYPLPVIV 135
QY 140 TMDMLGDKRDHEKISEW-----HSGVAD-----FITSISQSPPEARHSLWCS 182
D 136 IAEILGVPPADRTLFRSWSDRMLQMQVADPADMQFGDDADEYQRLVLPKPFMRMHA---- 191
```

QY 183 EQLSOYLMPIKERRVWPGSDLSILCTSEYEGMALSDKDLALILNVLAAETPADKTL 242
Db 192 -----YLHDHVTDRARPANDLISALVAARVGERLTDEQIVFEGALLMAGHVTSMLL 246
QY 243 ALMIYHLNPNQMDVLAADSLVPRATIAETIRYKPPVQLIPRQLSQDVTVVGMEIKKDT 302
Db 247 GNTVCLKDPRAEAAARADRSILIPALIEVLRPPITVMARVTTKOTVLAGTIPAGR 306
QY 303 IYFCMIGANRDPEAFQPDVNIHREDLGIKSAFSGAARHLAFSGIHCNCGAAPAKNE 362
Db 307 MVVPSLLSANHDEQVFTDPDHLDLARE-----GRQIAFGHIGHYICLGAFLARLE 355
QY 363 ISIVANIVLDKMRNIRLEDFC-----YAESGLY-----TRGP 395
Db 356 GRALAEALDFR-----PDFSPTGAKURHHRDGLFGVKNLPLTVRRGP 399

RESULT 7

US-09-335-409-8
; Sequence 8, Application US/09335409
; Patent No. 6121029

; GENERAL INFORMATION:

; APPLICANT: Schupp, Thomas

; APPLICANT: Ligon, James

; APPLICANT: Molnar, Istvan

; APPLICANT: Zirkle, Ross

; APPLICANT: Cyr, Devon

; APPLICANT: Goerlach, Joern

; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES

; FILE REFERENCE: 4-30582A

; CURRENT APPLICATION NUMBER: US/09/335,409

; CURRENT FILING DATE: 1999-06-17

; NUMBER OF SEQ ID NOS: 30

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 8

; LENGTH: 419

; TYPE: PRT

; ORGANISM: Sorangium cellulosum

US-09-335-409-8

Query Match 16.9%; Score 352.5; DB 3; Length 419;
Best Local Similarity 27.2%; Pred. No. 3.8e-29;
Matches 106; Conservative 85; Mismatches 170; Indels 29; Gaps 10;

QY 8 FSVLSDFQFQNNPYAFVSQLEEDPVHYEESIDSYFISRYHDVRYILOHDPDIFTTKSLVE- 66
Db 17 FKPFAPGYAEDPFAIERLEATPIFYWDEGRSWLTRYHDVSAVFRDRFAVSREEWES 76
QY 67 RAEPVMRGVLQAM--HG-----KEHSARRIVVRSPFGDLDHLSPLIKQNAENLLAP 118
Db 77 SAEYSSAIPELSDMKYGLGFLPPEDHARVRKLVNPSFTSRAIDLRLAEIQTVDQLLDA 136
QY 119 YLERGKSLVNDFGKTFVAVCTVMDMLGDKRDHEKISEWHSGVA-----DFITSISQSP 173
Db 137 RSGQBEFVVRDYAEGIPMRASALKVPACDEKFRFGSATARALGVLVPQVDEETK 196
QY 174 ARAHSLWCSEQLSQYLMPIKERRVNP--GSDLSILCTSEYEGMALSDKDLALILNVL 232
Db 197 TLVASV--TEGLA--LLHDVLDERRRNPENLVLTMLQAEADGSLSTKELVALVGAI 253
QY 233 AATEPADKTLALMIYHLNPNQMDVLAADSLVPRATIAETIRYKPPVQL--IPRQLSQDT 291
Db 254 AGTDTTIYLIADFVNLRLSPLEALVKAEPGLMRNALDEVLRFDNLRIGTVRFA 313
QY 292 VVGMEIKKDTTVFCMIGANRDPEAFQPDVNIHREDLGIKSAFSGAARHLAFSGI 351
Db 314 EYCGASIKKGMVFLIPSLARDGTVFSPDPVDFVRRD-----TGAS--LAYGRGP 363
QY 352 NCVGAAFAKNEIEIVANIVLDKMRNIRLEE 381
Db 364 VCPGVSLARLEAIEAVGTIFRRFPPEMKLKE 393

RESULT 9

US-09-568-102-8

; Sequence 8, Application US/09568102

; Patent No. 6346404

; GENERAL INFORMATION:

; APPLICANT: Schupp, Thomas

; APPLICANT: Ligon, James

; APPLICANT: Molnar, Istvan

```

; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-102-8

Query Match
Best Local Similarity 16.9%; Score 352.5; DB 3; Length 419;
Matches 106; Conservative 85; Mismatches 170; Indels 29; Gaps 10;

QY 8 FSVLSQFQNNPYAFYFQREEDPVHYEESIDSYFISRYHDVRYILOHDPDIETTKSLVE- 66
DB 17 FKPFAPGYAEDPFAIERLEATPIFYWDEGSRWLTTRYHDSAVFRDERFAVSREEWES 76
QY 67 RAEPVMRGVLAQM--HG-----KEHSKRRIVRSFSGDLDHLSPLIKQNAENLLAP 118
DB 77 SAEYSSAIPELSDMKYGLFGLPPEDHARVKLVNPSFTSRAIDLRLRAEIQTVDQLLDA 136
QY 119 YLERGKSLVNDFGKTFVAVCVTMDMLGLDKRDHEKISEWHSGVA-----DFITSISQSPE 173
DB 137 RSGQEEFVVDYAEIGIPMRAISALLKVPACDEKFRFRFGSATARALGVLGPVQVDEETK 196
QY 174 ARAHSLWCSEQLSQYLMPIVKERRVNP-GSDLSILCTSEYEGMALSDKDILAILINVL 232
DB 77 SAEYSSAIPELSDMKYGLFGLPPEDHARVKLVNPSFTSRAIDLRLRAEIQTVDQLLDA 136
QY 197 TLVASV--TEGLA--LLHDVLDERRNPLENDVLTMLQAEADGSRSLTKELVALVGAI 253
QY 233 AATEPADKTLALMIYHLNPNQMDVLDADSLVPRAIAETLRYKPPVOL-IPRQLSQDT 291
DB 254 AGTDTTIVLIAPAVLNLRSPALSLVKAEPGLMRNALDEVLRFDNILRIGTVRFARQDL 313
QY 292 VVGMGEIKKDTIVFCWIGMGAANDPEAFQPDVFNHREDLGIKSAFSGAARHLAFSGIH 351
DB 314 EYCGASIKKGMVFLIPLSALRDGTVFSRPDVFVRRD-----TGAS--LAYGRGPH 363
QY 352 NCVGAFAFNAEIEIVANIVLDKMRNIRLEE 381
DB 364 VCPGVSLARLEAEIAGVTIFRRFPPEMKLKE 393

RESULT 11
US-09-568-480-8
; Sequence 8, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-480-8

Query Match
Best Local Similarity 16.9%; Score 352.5; DB 3; Length 419;
Matches 106; Conservative 85; Mismatches 170; Indels 29; Gaps 10;

QY 8 FSVLSQFQNNPYAFYFQREEDPVHYEESIDSYFISRYHDVRYILOHDPDIETTKSLVE- 66
DB 17 FKPFAPGYAEDPFAIERLEATPIFYWDEGSRWLTTRYHDSAVFRDERFAVSREEWES 76
QY 67 RAEPVMRGVLAQM--HG-----KEHSKRRIVRSFSGDLDHLSPLIKQNAENLLAP 118
DB 77 SAEYSSAIPELSDMKYGLFGLPPEDHARVKLVNPSFTSRAIDLRLRAEIQTVDQLLDA 136
QY 119 YLERGKSLVNDFGKTFVAVCVTMDMLGLDKRDHEKISEWHSGVA-----DFITSISQSPE 173
DB 137 RSGQEEFVVDYAEIGIPMRAISALLKVPACDEKFRFRFGSATARALGVLGPVQVDEETK 196
QY 174 ARAHSLWCSEQLSQYLMPIVKERRVNP-GSDLSILCTSEYEGMALSDKDILAILINVL 232
DB 77 SAEYSSAIPELSDMKYGLFGLPPEDHARVKLVNPSFTSRAIDLRLRAEIQTVDQLLDA 136
QY 197 TLVASV--TEGLA--LLHDVLDERRNPLENDVLTMLQAEADGSRSLTKELVALVGAI 253
QY 233 AATEPADKTLALMIYHLNPNQMDVLDADSLVPRAIAETLRYKPPVOL-IPRQLSQDT 291
DB 254 AGTDTTIVLIAPAVLNLRSPALSLVKAEPGLMRNALDEVLRFDNILRIGTVRFARQDL 313
QY 292 VVGMGEIKKDTIVFCWIGMGAANDPEAFQPDVFNHREDLGIKSAFSGAARHLAFSGIH 351
DB 314 EYCGASIKKGMVFLIPLSALRDGTVFSRPDVFVRRD-----TGAS--LAYGRGPH 363
QY 352 NCVGAFAFNAEIEIVANIVLDKMRNIRLEE 381
DB 364 VCPGVSLARLEAEIAGVTIFRRFPPEMKLKE 393

RESULT 10
US-09-567-969-8
; Sequence 8, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
```

Db 137 RSGQEFVVDVYAEIGPMRAISALLKVPACDEKFRFGSATARALGVLVPQVDEETK 196
QY 174 ARAHSLWCSEQLSQYLMPIKERRVNP-GSDLSILCTSEYEGMALSDKDILALINVL 232
Db 197 TLVASV--TEGLA-LLHDVLDERRRNPLENDVLTMLLQAEADGSLSTKELVALVGAIIA 253
QY 233 AATEPADKTLALMIYHLNNPQMDVLDADRSIVPRAETLRYKPPVQL-IPRQLSQDT 291
Db 254 AGTDTTIIYLIAPAVLNLLSPALELVKAEPLMRNALDEVLFNLIIRIGTVRFARQDL 313
QY 292 VVGMEIKKDTIVFCMIGANRDPFAEQPDVFNHREDLGIKSAPSGAARHLAFSGGTH 351
Db 314 EYCGASIKKGEVFLIPSALEDGTVFSRPDVFVDRD-----TGAS--LAYGRGP 363
QY 352 NCVGAFAKNEIEIVANIVLDKMRNIRLEE 381
Db 364 VCPGVSLARLEAEIAGVTIFRRFPPEMKLKE 393

RESULT 12

US-09-568-486-8

; Sequence 8, Application US/09568486

; Patent No. 6355459

; GENERAL INFORMATION:

; APPLICANT: Schupp, Thomas

; APPLICANT: Ligon, James

; APPLICANT: Molnar, Istvan

; APPLICANT: Zirkle, Ross

; APPLICANT: Cyr, Devon

; APPLICANT: Goerlach, Joern

; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES

; FILE REFERENCE: 4-30582A

; CURRENT APPLICATION NUMBER: US/09/568,486

; PRIOR FILING DATE: 2000-05-10

; PRIOR APPLICATION NUMBER: 09/335,409

; PRIOR FILING DATE: 1999-06-17

; NUMBER OF SEQ ID NOS: 30

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 8

; LENGTH: 419

; TYPE: PRT

; ORGANISM: Sorangium cellulosum

US-09-568-486-8

Query Match 16.9%; Score 352.5; DB 3; Length 419;
Best Local Similarity 27.2%; Pred. No. 3.8e-29;
Matches 106; Conservative 85; Mismatches 170; Indels 29; Gaps 10;

QY 8 FSVLSDFQNNPYAFYFSQLRDEDPVHYEESIDSYFISRYHDVRYILOHPDITFTKSLVE- 66
Db 17 FKPFAPGYAEDPFAIERLEATPIFYWDGSRWLTTRYHDVSAVFRDERFAVSREWEES 76
QY 67 RAEPVNRGPVLAQM--HG-----KEHSKRRIIVRSFTIGDALDHLSPKIQNAENLLAP 118
Db 77 SAEYSSAIPELSDMKYKGLFGLPPEDHARVKLVNPSFTSRAIDLRLAEIQTVDQLLDA 136
QY 119 YLERGKSLVNDPFGKTFVAVCVTMDMLGDKRDHEKISEWHSGVA-----DFITSISQSP 173
Db 137 RSGQEFVVDVYAEIGPMRAISALLKVPACDEKFRFGSATARALGVLVPQVDEETK 196
QY 174 ARAHSLWCSEQLSQYLMPIKERRVNP-GSDLSILCTSEYEGMALSDKDILALINVL 232
Db 197 TLVASV--TEGLA-LLHDVLDERRRNPLENDVLTMLLQAEADGSLSTKELVALVGAIIA 253
QY 233 AATEPADKTLALMIYHLNNPQMDVLDADRSIVPRAETLRYKPPVQL-IPRQLSQDT 291
Db 254 AGTDTTIIYLIAPAVLNLLSPALELVKAEPLMRNALDEVLFNLIIRIGTVRFARQDL 313
QY 292 VVGMEIKKDTIVFCMIGANRDPFAEQPDVFNHREDLGIKSAPSGAARHLAFSGGTH 351
Db 314 EYCGASIKKGEVFLIPSALEDGTVFSRPDVFVDRD-----TGAS--LAYGRGP 363
QY 352 NCVGAFAKNEIEIVANIVLDKMRNIRLEE 381
Db 364 VCPGVSLARLEAEIAGVTIFRRFPPEMKLKE 393

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2004, 13:45:11 ; Search time 118.537 Seconds
(without alignments)
1220.359 Million cell updates/sec

Title: US-10-627-124-2
Perfect score: 2080
Sequence: 1 MSQSIKLSVLSDFQNNPY.....AESGLYTRGPVSLVAFDGA 405

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1585576 seqs, 357178320 residues

Total number of hits satisfying chosen parameters: 1585576

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	2080	100.0	405	15	US-10-627-124-2
2	454	21.8	410	14	Sequence 2, Appli
3	432	20.8	403	15	Sequence 6, Appli
4	430.5	20.7	404	14	Sequence 20, Appl
5	429.5	20.6	405	14	Sequence 16, Appl
6	423.5	20.4	404	14	Sequence 12073, A
7	413	19.9	395	14	Sequence 50, Appl
8	412.5	19.8	408	14	Sequence 12, Appl
9	406	19.5	392	14	Sequence 4, Appli
10	399.5	19.2	425	14	Sequence 32, Appl
11	398	19.1	388	14	Sequence 20, Appl
12	396.5	19.1	399	14	Sequence 13776, A
13	392.5	18.9	400	15	Sequence 8126, Ap
					Sequence 62072, A

14	391	18.8	418	14	US-10-132-134-40
15	389	18.7	393	14	US-10-156-761-9525
16	376	18.1	404	14	US-10-156-761-10431
17	374	18.0	430	14	US-10-145-415-95
18	373.5	18.0	430	14	US-10-145-415-30
19	371.5	17.9	430	14	US-10-145-415-4
20	371.5	17.9	430	14	US-10-145-415-28
21	370.5	17.8	457	14	US-10-156-761-11073
22	366.5	17.6	430	14	US-10-145-415-8
23	366.5	17.6	430	14	US-10-145-415-12
24	366.5	17.6	430	14	US-10-145-415-20
25	365.5	17.6	404	14	US-10-214-446-40
26	365	17.5	429	14	US-10-145-415-14
27	362.5	17.4	424	15	US-10-229-148B-13
28	361.5	17.4	404	14	US-10-321-188-76
29	361.5	17.4	415	14	US-10-214-446-56
30	361	17.4	414	14	US-10-156-761-13368
31	361	17.4	475	14	US-10-145-415-16
32	360	17.3	416	9	US-09-861-289-39
33	360	17.3	416	9	US-09-860-846-39
34	360	17.3	416	10	US-09-988-384B-39
35	360	17.3	416	10	US-09-836-821-39
36	360	17.3	416	10	US-09-793-708-18
37	360	17.3	416	14	US-10-201-365-13
38	360	17.3	416	14	US-10-160-539-18
39	360	17.3	416	14	US-10-271-889-39
40	356.5	17.1	418	15	US-10-461-194-118
41	356	17.1	475	14	US-10-145-415-22
42	353	17.0	430	9	US-09-738-626-4117
43	352.5	16.9	399	14	US-10-314-657-37
44	352.5	16.9	419	13	US-10-044-717-8
45	352.5	16.9	430	14	US-10-145-415-26

ALIGNMENTS

RESULT 1

US-10-627-124-2
; Sequence 2, Application US/10627124
; Publication No. US20040096944A1
; GENERAL INFORMATION:
; APPLICANT: NOVARTIS BIOTECH, INC.
; TITLE OF INVENTION: Methods For Producing Biological Substances In Pigment-Deficient
; FILE OF INVENTION: Mutants Of Bacillus Cells
; FILE REFERENCE: 10302.200-WO
; CURRENT APPLICATION NUMBER: US/10/627,124
; CURRENT FILING DATE: 2003-07-25
; PRIOR APPLICATION NUMBER: US 60/398,853
; PRIOR FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 405
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-10-627-124-2

APP/CONF

Query Match	100.0%;	Score 2080;	DB 15;	Length 405;
Best Local Similarity	100.0%;	Pred. No. 8.2e-187;		
Matches 405;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MSQSIKLSVLSDFQNNPYAFSOLREDDPVHYEESIDSYFISRYHDVRYILOHPDIPT	60	
Db	1	MSQSIKLSVLSDFQNNPYAFSOLREDDPVHYEESIDSYFISRYHDVRYILOHPDIPT	60	
QY	61	TKSLVERAEPVNRGPGVLAQMKGKHSKRRIIVRSGFDALDHLSPILKONAEENLAPYL	120	
Db	61	TKSLVERAEPVNRGPGVLAQMKGKHSKRRIIVRSGFDALDHLSPILKONAEENLAPYL	120	
QY	121	ERKGSDLVNDFOKTFACVCTMDMLGDKRDHKEISWHSGVADFTITSQSPEARASLW	180	
Db	121	ERKGSDLVNDFOKTFACVCTMDMLGDKRDHKEISWHSGVADFTITSQSPEARASLW	180	

QY 181 CSEQLSQYLMPIKERRVNPVSGDLISILCTSEYEGMALSDKDLALILNVLAAATEPADK 240
DB 181 CSEQLSQYLMPIKERRVNPVSGDLISILCTSEYEGMALSDKDLALILNVLAAATEPADK 240
QY 241 TLALMIYHLLNPNQNDVLAADRSVLPRAIAETLRYKPPVQLIPQLSQDVTVVGMEIKK 300
DB 241 TLALMIYHLLNPNQNDVLAADRSVLPRAIAETLRYKPPVQLIPQLSQDVTVVGMEIKK 300
QY 301 DTVFCMIGANRDPEAFQPPVFNHREDLIGIKAFSGAARHLAFGSGIHCNCVGAFAK 360
DB 301 DTVFCMIGANRDPEAFQPPVFNHREDLIGIKAFSGAARHLAFGSGIHCNCVGAFAK 360
QY 361 NEIEIVANVLDMKNIRLEEDFCVAESGLYTRGPVSLVAFDGA 405
DB 361 NEIEIVANVLDMKNIRLEEDFCVAESGLYTRGPVSLVAFDGA 405

RESULT 2
US-10-205-032-6
; Sequence 6, Application US/10205032
; Publication No. US20030113874A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; APPLICANT: Staffa, Alfredo
; TITLE OF INVENTION: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSARAMICIN
; FILE REFERENCE: 3016-2US
; CURRENT APPLICATION NUMBER: US/10/205,032
; CURRENT FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 410
; TYPE: PRT
; ORGANISM: micromonospora carbonacea subspecies aurantiaca
US-10-205-032-6

Query Match 21.8%; Score 454; DB 14; Length 410;
Best Local Similarity 29.3%; Pred. No. 9.8e-34;
Matches 117; Conservative 84; Mismatches 156; Indels 42; Gaps 9;
QY 22 YFSQLEEDPVHYEESIDSYFISRYHDVRIYLQHPDI----FTTKSLVERABPVMRGP-V 76
DB 26 WFAYMRKNWPVSWDETRQAWHVFSDYQVTNPLIFSSDFTSVFPVPSLALLMGECT 85
QY 77 LAQMKGKHSKRIRVRSFIDGDLHLSPLIKQNAENLLAPYLRGKSDLVNDFGKTF 136
DB 86 IGGIDPPRHAPRLKLVQAFTPRRIAQMELRIGQITADVLDQVRDQDRIDIASDLAYPLP 145
QY 137 VCVTMDMLGDKRDHEKISEW-----HSGVA-----DFITSISQSPEARAHSLWCSEQ 184
DB 146 VTVIAELGIPDKHEKFEWVDIILNSGLEYPNLPDDFTETVGFA-----IEE 195
QY 185 LSQYLMPIKERRVNPVSGDLISILCTSEYEGMALSDKDLALILNVLAAATEPADKTLAL 244
DB 196 WSEFLYAQIAHKAEPKODLISGLCAAEDVGKLTDEEVNIVA-LLTAGHISATLLS 254
QY 245 MYHLL-NNPEQNDVLAADRSVLPRAIAETLRYKPPVQLIPQLSQDVTVVGMEIKKXTI 303
DB 255 NLFLVLEHPQAAAVRADRSVLPVGIETLRYSGPFCIFRILNEDTDLGHMPRKGM 314
QY 304 VFCMIGANRDPEAFQPPVFNHREDLIGIKAFSGAARHLAFGSGIHCNCVGAFAKNEI 363
DB 315 VTAWASANRDTEVFTDPTFDIRE-----SNKHLAFGHIHCLGAFLAFLAUEA 364
QY 364 EIVANIVLDMKNIRLEE---DFCAESGLYTRGPVSL 399
DB 365 KVFLNQTLQDFTFRIDHVGVEFYDADQLTARLPPQVV 403

RESULT 3
US-10-229-148B-20

; Sequence 20, Application US/10229148B
; Publication No. US20040091975A1
; GENERAL INFORMATION:
; APPLICANT: Meiji Seika Kaisha, Ltd.
; TITLE OF INVENTION: Midecamycin biosynthetic genes
; FILE REFERENCE: 138451 US
; CURRENT APPLICATION NUMBER: US/10/229,148B
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: 210516/2002
; PRIOR FILING DATE: 2002-07-19
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Streptomyces mycarofaciens
US-10-229-148B-20

Query Match 20.8%; Score 432; DB 15; Length 403;
Best Local Similarity 30.0%; Pred. No. 1.1e-31;
Matches 111; Conservative 76; Mismatches 159; Indels 24; Gaps 6;
QY 22 YFSQLEEDPVHYEESIDSYFISRYHDVRIYLQHPDIFTTKSLVERAEP-----VMRGP-V 76
DB 23 WFAFNRTTHPWFWDSESHAWQVFRYDDYLTVSNNPQFFSSDFNEVMPTPELEWVGPGT 82
QY 77 LAQMKGKHSKRIRVRSFIDGDLHLSPLIKQNAENLLAPYLRGKSDLVNDFGKTF 136
DB 83 IGALDPPAHGPMRKLVSQAFTPRRVARLEPRIRAVTOELLDVARGOETIDVVGGDLSYALP 142
QY 137 VCVTMDMLGDKRDHEKISEW-----HSGVADFTITSQSPEARAHSLWCSEQLSQYLM 190
DB 143 VIVIAELGIPSGDRDVRGWDVTLITNEGL-EYNLPDNFSETIAPAL--KEMTDYLL 198
QY 191 PVIKERRVNPVSGDLISILCTSEYEGMALSDKDLALILNVLAAATEPADKTLALMIYHLL 250
DB 199 HQIHAKREAPVDDLTISGLVQAEDGKRLTDVEIVNIVALLTAGHVSSSTLLSNLFLVLE 258
QY 251 NNPEQNDVLAADRSVLPRAIAETLRYKPPVQLIPQLSQDVTVVGMEIKKDTIVFCMICA 310
DB 259 ENPQALADLRADRELVTGAVEETLRYSPFNINFRFLKEDTDLGPEMKKGQVIAWSQS 318
QY 311 ANRDPEAFQPPVFNHREDLIGIKAFSGAARHLAFGSGIHCNCVGAFAKNEIEIVANIV 370
DB 319 ANRDPEHPEPDTDIR-----SSSRHMAFGIGIHCCLGAFLAQRGKVVLELM 369
QY 371 LDKMRNIRLE 380
DB 370 LDQVREFRID 379

RESULT 4
US-10-214-446-16
; Sequence 16, Application US/10214446
; Publication No. US20030180742A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, David
; APPLICANT: Burk, Mark J.
; APPLICANT: Hitchman, Tim
; APPLICANT: Pujol, Catherine
; APPLICANT: Richardson, Toby
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING
; TITLE OF INVENTION: THEM AND METHODS OF MAKING AND USING THEM
; FILE REFERENCE: 09010-500001
; CURRENT APPLICATION NUMBER: US/10/214,446
; CURRENT FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/309,497
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 404

[illegible]

QY 289 QDTVGGMEIKKDTIVFCMIGANRDPEAFQPDVFNHREDLGKSAFSGAARHLAFGS 348
Db 296 EDFEGGAEBVRGDLVLMGANDRPEAFDDPTDFDLTRNFTG-----HLRQGW 345
QY 349 GHNCVGAFAFAKNEIEIVANIVLDKMRNRL 379
Db 346 GPHVCVGAALAEQVSRFTLLDRLPGLEL 376

RESULT 7

US-10-214-446-12
; Sequence 12, Application US/10214446
; Publication No. US20030180742A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, David
; APPLICANT: Burk, Mark J.
; APPLICANT: Hitchman, Tim
; APPLICANT: Pujol, Catherine
; APPLICANT: Richardson, Toby
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING
; FILE REFERENCE: 09010-500001
; CURRENT APPLICATION NUMBER: US/10/214,446
; CURRENT FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/309,497
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-214-446-12

Query Match 19.9%; Score 413; DB 14; Length 395;
Best Local Similarity 28.6%; Pred. No. 6.7e-30;
Matches 112; Conservative 72; Mismatches 162; Indels 46; Gaps 8;
QY 8 FSVLSDPQNNPVAYFSQLEEDPVHYEESIDSYFSRYHDVRYILQHDPDIFTT-----61
Db 5 FNPYCEEFTQNPQNFALRTQDPVHYIEFPANALFGFEDVWRAGMDRESTATYGGSP 64
QY 62 -KSLVERAEPVMRGPVLAQMKGKHSKRRIVVRSPFIGDALDHLSPLIKQNAENLLAPYL 120
Db 65 QALLIDR---VKQPEIFLFWMDIPNMIHRIIAKDYGRNAMPLEKIRATAKEAITPYL 121
QY 121 ERGKSLVNDPFGKTFACVVTMDMLGDKRDHEKI-----SEWHSGVADFTITSIQ 170
Db 122 KSGEMD-VYAFARTVALFTIADMIGLRPEEVVRIIRSLIDIFFGRTPGHRT-----171
QY 171 SPEARAHSLWCSEQLSQYLMPIVKERRVN--PGSDLIISILCTSEYEGMALSDKDILALIL 228
Db 172 TPDGVA----AFHEVAYVLDLIGHYRAKAGPEGSHIDNWLKAEPPDGRPLDDQALCANIF 227
QY 229 NVLLAATEPADKTLALMIYHLNPNQMDVLAADRSVLPRAIATLRYKPPVQLIPRQLS 288
Db 228 SLSITGSDTVPLSSAAIYYLSEHPAQLEAVRSRALIPAAFAETVRYDQPTNVLRGLLA 287
QY 289 QDTVGGMEIKKDTIVFCMIGANRDPEAFQPDVFNHREDLGKSAFSGAARHLAFGS 348
Db 288 IDTDKYGKPMKQVQALFMYASANRDLPEFHPDPTENIYRD-----PRRTLSFGS 337
QY 349 GHNCVGAFAFAKNEIEIVANIVLDKMRNRL 380
Db 338 GHICIGQLLAKLEGGQILLETLEFHIPTDFTVQ 369

RESULT 8

US-10-214-446-4

; Sequence 4, Application US/10214446
; Publication No. US20030180742A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, David
; APPLICANT: Burk, Mark J.
; APPLICANT: Hitchman, Tim
; APPLICANT: Pujol, Catherine
; APPLICANT: Richardson, Toby
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING
; FILE REFERENCE: 09010-500001
; CURRENT APPLICATION NUMBER: US/10/214,446
; CURRENT FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/309,497
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 408
; TYPE: PRT
; ORGANISM: Bacterial
US-10-214-446-4
Query Match 19.8%; Score 412.5; DB 14; Length 408;
Best Local Similarity 26.6%; Pred. No. 7.8e-30;
Matches 109; Conservative 88; Mismatches 166; Indels 47; Gaps 10;
QY 16 QNNPYAFSQRLEE---DPVHYEESIDSYFSRYHDVRYILQHDPDIFTTKSLV-----65
Db 19 QNPYPFLFERIREHGVQVRVRLNPTLEVMMVTGYDEAVAALTDPRLSSPSPGVNGLEEM 78
QY 66 ---ERAEPVMRGPVLAQMKGKHSKRRIVVRSPFIGDALDHLSPLIKQNAENLLAPYL 122
Db 79 AHQERTNVLMSMLVA--NGEDHTLRNLVSKAFTARRVEQLAPRQVQAHDTDAFLDAVAAR 136
QY 123 GKSLVNDPFGKTFACVVTMDMLGDKRDHEKISEWHSGVADF-----ITSISQSPPEARA 176
Db 137 GSADLVSEFALPLPNAVLSDLIGIPA-----EGQDPFARLAVGLIMPNTPERLA 186
QY 177 HSLWCSEQLSQYLMPIVKERRVNPQSDLIISILCTSEYEGMALSDKDILALINVLAAATE 236
Db 187 KGARARAEELTFEPFLIAQKKEPKDILLALCAAQAE-ERISDRELTAMTILLTAGHE 245
QY 237 PADKTLALMIYHLNPNQMDVLAADRSVLPRAIATLRYKPPVQL-IPRQLSQDTPVVG 295
Db 246 TTASLIANGVHALLRHPQFATLRDDPSLLPGAIBELLRYEGPVSRGVARFTTDPYEIGG 305
QY 296 MEIKKDTIVFCMIGANRDPEAFQPDVFNHREDLGKSAFSGAARHLAFSGIHNCVG 355
Db 306 VTVPAGEMIITGLAANRDPRYDRPDILDVAREV-----PQQLAFGHGVHFCJG 356
QY 356 AAFAKNEIEIVANIVLDKMRNRL---BEDFCYAESGLYTRGFVSLVAF 402
Db 357 APLARAARIAIGTLRRFPDLRLADPDADLSRREGIL--RGWATLPVTF 404
RESULT 9
US-10-214-446-32
; Sequence 32, Application US/10214446
; Publication No. US20030180742A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, David
; APPLICANT: Burk, Mark J.
; APPLICANT: Hitchman, Tim
; APPLICANT: Pujol, Catherine
; APPLICANT: Richardson, Toby
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING
; FILE REFERENCE: 09010-500001
; CURRENT APPLICATION NUMBER: US/10/214,446
; CURRENT FILING DATE: 2002-08-05

```
; PRIOR APPLICATION NUMBER: US 60/309,497
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Bacterial
; US-10-214-446-32

Query Match
Best Local Similarity 19.5%; Score 406; DB 14; Length 392;
Matches 113; Conservative 65; Mismatches 172; Indels 26; Gaps 8;

QY 15 FQNNPYAYFSQLEEDPVHESIDSY---FISRYHDVRYIQLQHPDIFTTKSL--VERAE 69
DB 13 FTANPYPYAKLREAGVHVVRMPDGFQFWLVVVGHEGRAALADPRLAKSPSIVGVRPE 72
QY 70 PVMRGVLAQMKGKHSKRRIIVRSFIGDALDHLSPKQNAENLLAPYLRGKSDLYN 129
DB 73 EDIIGVHLLAADAPDHTLRRLVGTGTRRVEGLRPRIQQLTTELADAMEPAGRADLYD 132
QY 130 DFGKTFVAVCVTMDMLGDKRDKEKISEHSGVADFTISQSPPEARAHSLWCSEQLSQYL 189
DB 133 AFAYPLPIVICELGVPADRDFTFRNSN---QLVPTGDDQFQGA-----WVDFAYL 184
QY 190 MPVIKERR-VNPGSDILISICTSEYV-GMALSDKOILALINLVLAATEBPADKTLALMTY 247
DB 185 DALIEDKRAAGPTDILLSALITARAEDGRLSGPELRANAYLLIAGHETTVNLIANTVR 244
QY 248 HLLNPEQNDVLAORSVLPRAIAETLYRKYPPVQLIPQLSQDQTV-VGMEIKKOTIVFC 306
DB 245 NLLTHPEQLAALRADPDLDDGTIESLRYDGPVETGTFRTREAVTIGGREIAAGQYVLV 304
QY 307 MIGAARDPEAFQPDVFNHREDLGKSAFSGAARHLAGSGIHCNVCVGAFAKNEIEIV 366
DB 305 GIGALDRDPDPDPDFDIRDTRG-----HLAFGHIHYCLGAPLARLEGRIA 354
QY 367 ANIVLDKMNIRLEED 382
DB 355 LRTLLDRFPDLELDPE 370

RESULT 10
US-10-214-446-20
; Sequence 20, Application US/10214446
; Publication No. US20030180742A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, David
; APPLICANT: Burk, Mark J.
; APPLICANT: Hitchman, Tim
; APPLICANT: Pujol, Catherine
; APPLICANT: Richardson, Toby
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING
; TITLE OF INVENTION: THEM AND METHODS OF MAKING AND USING THEM
; FILE REFERENCE: 09010-500001
; CURRENT APPLICATION NUMBER: US/10/214,446
; CURRENT FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/309,497
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 425
; TYPE: PRT
; ORGANISM: Bacterial
; US-10-214-446-20

Query Match
Best Local Similarity 19.2%; Score 399.5; DB 14; Length 425;
Matches 113; Conservative 82; Mismatches 163; Indels 69; Gaps 12;

; PRIOR APPLICATION NUMBER: US 60/309,497
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Bacterial
; US-10-214-446-32

Query Match
Best Local Similarity 30.1%; Pred. No. 3e-29;
Matches 113; Conservative 65; Mismatches 172; Indels 26; Gaps 8;

QY 14 QFQNNPYAYFSQLEEDPVH-----YBESIDSYFISRYHDVRYIQLQHPDIFTTKSLVERAE- 69
DB 23 EFASDPYPAYAWLRSHAPVHRTTLP SGVEAWLVTRYGDARQAL--ADQRLSKNPAHDES 80
QY 70 PVMRGV-----LAQMKGKHSKRRIIVRSFIGDALDHLSPKQNAENLL 116
DB 81 PHAKGTGIPGERKAEMLTHLNIIDPPDHTLRRLVSKAFTPRVAEFTPRVQELTDRLI 140
QY 117 APYLRGKSDLYNDFGKTFVAVCVTMDMLGDKRDKEKISEW-----HSGVADFTI 165
DB 141 DAFVTKGSADLIHDAFPLPIVAICDLGVPEDQDDPRDWAGMMIRHGGGPRGGVARSV 200
QY 166 TSISQSPPEARAHSLWCSEQLSQYLMPVIKERRVNPVGSDDLISICTSEYEGMALSDKILA 225
DB 201 -----KMGYLAELIHRKEAPGDDLIISGLIKASDRGHEHLTENEAAA 243
QY 226 LILNVLAAATEPADKTLALMIYHLLNPN--BOMNDVLA--DRSLVPRRAIAETLYRKYPPV 280
DB 244 MAFILLFAGFETVNLGNVTVQLLRHFGQORRELQTSLAAGETGLLETGIEELLRYDGPV 303
QY 281 QLIP-RQLSQDQTVVGMEIKKOTIVFCMIGAANRDPEAFQPDVFNHREDLGKISAFSG 339
DB 304 EMATWRYATEPLTIGQDIPAGDPVLVLAADRDPERFDRPDVLDLARRD----- 354
QY 340 AARHLAFSGIHCNVCVGAFAKNEIEIVANIVLDKMNIRLEEDFCYAE-----SGLYTRGP 395
DB 355 -NQHLGYGHIHYCLGAPLARLEGOTATLTLTRLPDLRLAAD--PAELWRGGLIMRGL 411
QY 396 VSLVAVF 402
DB 412 RTLPEVF 418

RESULT 11
US-10-156-761-13776
; Sequence 13776, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13776
; LENGTH: 388
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
; US-10-156-761-13776

Query Match
Best Local Similarity 19.1%; Score 398; DB 14; Length 388;
Matches 110; Conservative 70; Mismatches 189; Indels 26; Gaps 7;

QY 18 NPYAYFSQLEEDPVHESID---SYFISRYHDVRYIQLQHPDIFTTK-----SLVERA 68
DB 10 DPAVYVDRLDTPAPVHRIAGTDGKPAWLVTRYDDVREGLANPULSLDKKHALPGNRYGLA 69
QY 69 EPVMRGPVLAQMKGKHSKRRIIVRSFIGDALDHLSPKQNAENLLAPYLRGKSDLY 128
DB 70 LPALDANLNDAPDHTLRRLVGAFTLRVVEQLREPVRETAHRLLDALGTHGSTDLI 129
QY 129 NDFGKTFVAVCVTMDMLGDKRDKEKISEHSGVADFTISQSPPEARAHSLWCSEQLSQY 188
```

Db 130 ASYAAPLPITVICDLGVDPDEHRRDFRAW---TDLVTPDPARPDPVARESV---VSLLGPF 183
QY 189 LMPVTKERRVNPGLISILCTSEYEGMALSOKOILALINLVLLAATEPADKTLALMIYH 248
Db 184 FTGLADKRKNPADLLSLIAVQBEGRDLTDELMSLAFLIFAGYENTVHLIGNAVLA 243
QY 249 LNNPEQNDVLADSLVPRATAETLRYPKPPVQLIPRQLS-ODTVVGGMEIKKOTIVFCM 307
Db 244 LURHEQALAEEDPARLPDVAFEYEGPALLAIRPPVRDVTGIGVTVPAGETVLLS 303
QY 308 IGAANRDPAPFPQPVFNTHREDLGIKSAFSGAARHLAFSGIHNCVGAFAFKNIEIIVA 367
Db 304 LSAANRDPFRDPD-----RLDLG-----RDAAGHLALGHGVHCLGAPLARLETEVAL 353
QY 368 NIVLDMNRNIRLEEDFCYAESGLYTRGPVSLIVAF 402
Db 354 AALLERFPDLAAETEPRRRPSLRAGLLALPVTY 388

RESULT 12

US-10-156-761-8126
; Sequence 8126, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8126
; LENGTH: 399
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8126

Query Match 19.1%; Score 396.5; DB 14; Length 399;
Best Local Similarity 28.4%; Pred. No. 2.4e-28;
Matches 111; Conservative 62; Mismatches 189; Indels 29; Gaps 7;
QY 18 NPAYVFSQLEEDPVHYEESIDSYFISRYHVDVRYILOHPDIFTTKSLVERAEP-----VNR 73
Db 18 DPYPLVLAELREAGPV-VRQADGSYLIGTYHEIVALLHDPMSADPR--SRTAPAPVEVTR 74
QY 74 GPVLAQMGKEHSKRIRVVRF-----IGDALHLSPLIKQNAENLLAPYLRGKSDLVN 129
Db 75 KPSFLRLDDPEHRLRNSAMPFGPHSPGRVDSMRGEIVQLTKELAEVFGREGRIQDVVD 134
QY 130 DFGKTFVAVCTVMDLGLDKRDHEKISEWHSGVADFTTISQSPPEARAHSLS-----WCSEQ 184
Db 135 DFAYPLPVTVICLLGIPKDKQLQFDW---TDTLVASADIGPEGDTAERDQAQAQOE 191
QY 185 LSQYLMPIVKERRVNPGLSILCTSEYEGMALSDKDIALLINLVLLAATEPADKTLAL 244
Db 192 MGQYLVQLAEQRGRGTGMDLSDLVNPEPPAARLSEEDLAANTILLFIAGHETTNNLIAN 251
QY 245 MYHLLNPNQNDVLADSLVPRATAETLRYPKPPVQLIPRQLSQDTPVGGMEIKKOTIV 304
Db 252 GVLTLRLRPDQLDRREDSLLPRAVEELLRYEPPVHMRERVPLDIDVAGTTIPGTSV 311
QY 305 FCWIGAANDPPEAFQEPDVFNTIHREDLGKSAFSGAARHLAFSGIHNCVGAFAFKNIEI 364

Db 312 ILALASGSRDPMRFSEPDPRDFDTPRD-----NQHVFGSGIHLFCGAPLARIEAE 361
QY 365 IVANIVLDMNRNIRLEEDFCYAESGLYTRGP 395
Db 362 AALGALLPHLGLTARLVQDPPPPYRQNAMLRGF 392
RESULT 13
US-10-282-122A-62072
; Sequence 62072, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62072
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Mycobacterium avium
US-10-282-122A-62072

Query Match 18.9%; Score 392.5; DB 15; Length 400;
Best Local Similarity 26.9%; Pred. No. 5.8e-28;
Matches 108; Conservative 84; Mismatches 173; Indels 37; Gaps 8;
QY 18 NPAYVFSQLEEDPV-----HYEESIDS-----YFISRYHVDVRYILOHPDIFTTKSLV 65
Db 18 DPYPLFARRRAGVAGTVMDSKTPMKQYSAVSFAVNTVFRDGRVFSKPY 76
QY 66 ERAEPVMRGPVLAQMGKEHSKRIRVVRSFITCDALDHLSP-LIKQNAENLLAPYLRGK 124
Db 77 DKTIGLFWGPTTLAMEGKKGRDHRNLVSAFSAKALARWEPTIVRPICNALIDDFDAGT 136
QY 125 DVLVNDFGKTFVAVCTVMDLGLDKRDHEKISEWHSGVADFTT---SISQSPPEARAHSLS 181
Db 137 ADLVQRQTFEFPTRVIRLILGLPDED---LPMFHTRAVOLISYHVDYERAFEA----- 186

182	QY	SQLSOYLMPIVKERRVNGSDLSIICTSBYEGWALSDXDIILALINVLVLAATEPADXT	241
187	Db	SAALKDYFLEQIQEOKSPTDIIIGDLVTAEIDGKLSDEAIYSFTRLLLPAGLETTERS	246
242	QY	LALMIYTHLLNPEQMNDVLADRSIVPRAIAETRLYKPPVQLIPIQLSODTVGVGMEIKD	301
247	Db	SGNLIYLLTHPDQFALQADRELLAPABEGLRFETPLTVVQRFTEDETLHGVRIPAR	306
302	QY	TIVFCWIGAAANDPEAFQPDQVFNHREDLGIKSAFSGAARHLAFSGHNCVGAFAFN	361
307	Db	SVIGVCIGSANDRRERWSESEDFIRKHV-----PHISFAAGHTCIGLHLARL	356
362	QY	EIEIVANIVLQKRNIRLEEDFCVAESGLYTRGPVSLLVAFD	403
357	Db	ETRFVAMECLNRLTNVTLSDGDPHIHGQPRSPALPTVFD	398

RESULT 14

US-10-132-134-40
; Sequence 40, Application US/10132134
; Publication No. US20030171562A1

```

: GENERAL INFORMATION:
: APPLICANT: Farnet, Chris
: APPLICANT: Yang, Xianshu
: APPLICANT: Staffa, Alfredo
: APPLICANT: Zazopoulos, Emmanuel
: TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES
: FILE REFERENCES: 3012-2US
: CURRENT APPLICATION NUMBER: US/10/132,134
: CURRENT FILING DATE: 2002-04-26
:

```

Query Match	18.8%	Score 391;	DB 14;	Length 418;
Best Local Similarity	26.7%;	Pred. No. 8.6e-28;		
Matches 111;	Conservative .84;	Mismatches 166;	Indels 54;	Gaps 11;

	QY	10	VLSDOFQN-----NPAYFSQLREEDPVHYEESIDSIFYSRHYDVRYYILQHDPDIFTKSL	64
	Db	17	I VTDVFQTEGKNPYPPLYRRLQELGQVHRSEQL-GWVATGYEVCSAALDRPRVKGPQE	75
	QY	65	VE-----RAEPVMRGVPVLAWHQHKEHSAKRIIVRSFIGDALDHLSPLIKQAENL	115
	Db	76	TOPGRPDPAHSAEAALLUG-TMHRLDDPDTRLRLLVNGAFTRPSVAALEPPDIQELIDDL	134
	QY	116	LAPYLERGS---DLVNDFGKTFAVCVTMDMLGLDKRDHEKISEWHSGVADPFITS---	168
	Db	135	ITPAVKKAEEGEVDMSGFAPPISAVIGMLGVPASDMHR---PHDVVLDSNVELG	191
	QY	169	----SOSPEARAHSLWCSEQLSOYLMPIWKERRNPNVGSDLIISIICTEYEGMALSKDILA	225
	Db	192	FTGEDLPKADA----AADELIATYFKLGAERMNRPADLTSTLANATEAGDRLTEQELVT	247
	QY	226	LILANLLAATEPADKTLALMIYHLINNPQONDVLADRSLVPRAIETLYRXPPOLI--	283
	Db	248	MULIFPMAGFETHTHSMGNMGFPALLEPEQTOWLRNNMDAMPAAEVELIRYSPOVFIAG	307
	QY	284	----PRQLSQDTVVGMGBIIKKDITTVFCMIGAANDRPDAEQPOPVFNHREDLGIKSAFE	339
	Db	308	YTKEPVELADGTAV----PADEVFLMIGANNEDPRVFDPELLRLDR-----G	352
	QY	340	AARHLAFGSGIHNCVGAAPAKNEIEIVANIVLDKQRNIRLEEDECFAEGLYTRG	394
	Db	353	EAPMSFGGIHCYGAGLALEIRKI FTSLTFRSAIELAEPEBRRSSGLARG	407

RESULT 15

US-10-156-761-9525

```

/ Sequence 9525, Application US/10156761
/ Publication No. US20030119018A1
/ GENERAL INFORMATION:
/ APPLICANT: OMURA, SATOSHI
/ APPLICANT: IKEDA, HARUO
/ APPLICANT: SHIKAWA, JUN
/ APPLICANT: HORIKAWA, HIROSHI
/ APPLICANT: SHIBA, TADAYOSHI
/ APPLICANT: SAKAKI, YOSHIYUKI
/ APPLICANT: HATTORI, MASAHIRA
/ TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
/ FILE REFERENCE: 249-762
/ CURRENT APPLICATION NUMBER: US/10/156,761
/ CURRENT FILING DATE: 2002-05-29
/ PRIOR APPLICATION NUMBER: JP 2001-204089
/ PRIOR FILING DATE: 2001-05-30
/ PRIOR APPLICATION NUMBER: JP 2001-272697
/ PRIOR FILING DATE: 2001-08-02
/ NUMBER OF SEQ ID NOS: 15109
/ SEQ ID NO 9525
/ LENGTH: 393
/ TYPE: PRT
/ ORGANISM: Streptomyces avermitilis
/ US-10-156-761-9525

```

[illegible]

Search completed: December 10, 2004, 14:02:29
Job time : 123.537 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2004, 11:14:31 ; Search time 417.058 Seconds
(without alignments)
9477.854 Million cell updates/sec

Title: US-10-627-124-7
Perfect score: 753
Sequence: 1 gtaagagatgacggaat.....aacggatcacatctgaagga 753

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_23Sep04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001as:*
- 5: Geneseq2001bs:*
- 6: Geneseq2002as:*
- 7: Geneseq2002bs:*
- 8: Geneseq2003as:*
- 9: Geneseq2003bs:*
- 10: Geneseq2003cs:*
- 11: Geneseq2003ds:*
- 12: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	753	100.0	753	12	ADJ62044
2	61.6	8.2	705	10	ACF70114
3	61.6	8.2	110000	10	ACF65385_2
4	61.6	8.2	110000	10	ACF67367_31
5	58.4	7.8	651	10	ACF66408
6	40.4	5.4	2000	8	ADA71938
7	39.8	5.3	2000	8	ADA71938
8	38	5.0	1194	8	ACA34459
9	38	5.0	110000	2	AAT42063_15
10	36.8	4.9	110000	10	ADH10017_0
11	36.8	4.9	110000	10	ADH10017_1
12	36.4	4.8	3048	2	AAQ75343
13	36.4	4.8	3211	2	AAQ75344
14	36.2	4.8	110000	6	ABA03041_24
15	36	4.8	1338	10	ACF72025
16	36	4.8	5580	4	ABL15162
17	36	4.8	110000	10	ACF67367_52
18	36	4.8	110000	10	ACF65387_3
19	35.4	4.7	736	4	AAH99421
20	35.4	4.7	825	4	ABL04657
21	35.4	4.7	1695	4	ABL20607

ALIGNMENTS

RESULT 1
ADJ62044
ID ADJ62044 standard; DNA; 753 BP.
XX
AC ADJ62044;
DT 06-MAY-2004 (first entry)
XX
DE Bacillus subtilis ymc encoding DNA SEQ ID NO:7.
XX
KW heterologous biological substance; Bacillus; ymc; red pigment; gene; ds.
OS Bacillus subtilis.
FH Key Location/Qualifiers
FT CDS 1..753
FT /*tag= a
FT /product= "ymc"

WO2004011609-A2.

05-FEB-2004.

25-JUL-2003; 2003WO-US023398.

26-JUL-2002; 2002US-0398853P.

(NOVO) NOVOZYMES BIOTECH INC.

Tang M, Sloma A, Sternberg D, Behr R;

WPI; 2004-143839/14.

P-PSDB; ADJ62045.

Producing a heterologous biological substance comprises cultivating pigment-deficient mutants of Bacillus cell that directs synthesis of the heterologous biological substance and has a modification of the cypX and ymc genes.

Example 1; SEQ ID NO 7; 62pp; English.

The present invention describes a method for producing a heterologous biological substance comprising cultivating a mutant of a parent Bacillus cell in a medium suitable for the production of a heterologous biological

Copyright

CC substance, and recovering the heterologous biological substance from the
CC cultivation medium, where the mutant cell comprises a first nucleic acid
CC sequence directing synthesis of the heterologous biological substance and
CC a second nucleic acid sequence comprising a modification of at least one
CC of the genes *cypX* and *ymcC*, which are involved in the production of a red
CC pigment. The mutant cell is also deficient in the production of the red
CC pigment compared to the parent *Bacillus* cell when cultivated under the
CC same conditions. Also described: (1) a mutant of a parent *Bacillus* cell,
CC comprising a first nucleic acid sequence directing synthesis of a
CC heterologous biological substance and a second nucleic acid sequence
CC comprising a modification of at least one of the genes *cypX* and *ymcC*,
CC which are involved in the production of a red pigment, wherein the mutant
CC cell is deficient in the production of the red pigment compared to the
CC parent *Bacillus* cell when cultivated under the same conditions; and (2)
CC obtaining a mutant of a parent *Bacillus* cell, comprising introducing into
CC the parent *Bacillus* cell a first nucleic acid sequence directing
CC synthesis of a heterologous biological substance and a second nucleic
CC acid sequence comprising a modification of at least one of the genes *cypX*
CC and *ymcC*, which are involved in the production of a red pigment, and
CC identifying the mutant cell comprising the modified nucleic acid
CC sequence, where the mutant cell is deficient in the production of the red
CC pigment compared to the parent *Bacillus* cell when cultivated under the
CC same conditions. The methods and compositions of the present invention
CC are useful for producing biological substances, e.g. nucleic acid,
CC polyamide, polypeptide, polypeptide, or polysaccharide, in pigment-
CC deficient mutants of *Bacillus* cells, and obtaining the pigment-deficient
CC mutant cells. The present sequence encodes *Bacillus* subtilis *ymcC*, which
CC is used in the exemplification of the present invention.
XX
SQ

Query Match 100.0%; Score 753; DB 12; Length 753;
Best Local Similarity 100.0%; Pred. NO. 2.3e-226;
Matches 753; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGAATGAGATGACCGGAATGTTAAACGGAAGAGGCTCTGTGCAATTTATTGCTGAGGCA 60
DB 1 GTGATGAGATGACCGGAATGTTAAACGGAAGAGGCTCTGTGCAATTTATTGCTGAGGCA 60
QY 61 TTAACAGAAACTGACAGAGAAATATTTGAACCGCGCAGGCGATGTTTGGTGGGATCAGC 120
DB 61 TTAACAGAAACTGACAGAGAAATATTTGAACCGCGCAGGCGATGTTTGGTGGGATCAGC 120
QY 121 CCATTTAAACGAGGTTTTCAGAGGATTTATTTACAGATTAATTTGATGGCGGAAGCT 180
DB 121 CCATTTAAACGAGGTTTTCAGAGGATTTATTTACAGATTAATTTGATGGCGGAAGCT 180
QY 181 CAATTTAAACGAGGTTTTCAGAGGATTTATTTACAGATTAATTTGATGGCGGAAGCT 240
DB 181 CAATTTAAACGAGGTTTTCAGAGGATTTATTTACAGATTAATTTGATGGCGGAAGCT 240
QY 241 CTTGGAACTCCGAGAGAAAGGCTGAAACGAAAGTAAGGAAAGGATATCACGAAACAGG 300
DB 241 CTTGGAACTCCGAGAGAAAGGCTGAAACGAAAGTAAGGAAAGGATATCACGAAACAGG 300
QY 301 AGATTTGAGAGAGAGCCCTTGTGGCTCATGCGCGGGATCCGAGGCGATTCATACATTT 360
DB 301 AGATTTGAGAGAGAGCCCTTGTGGCTCATGCGCGGGATCCGAGGCGATTCATACATTT 360
QY 361 TCTGATTTATAGATAACAAGCCTACAGCTGTTGAGACAAGAGTGTGAACATGCAATTT 420
DB 361 TCTGATTTATAGATAACAAGCCTACAGCTGTTGAGACAAGAGTGTGAACATGCAATTT 420
QY 421 TTTGAGAGGCTCATTTTTCAGATGCTGTTGTTGAGATGCTCTGTAAGCGATTAATCGGG 480
DB 421 TTTGAGAGGCTCATTTTTCAGATGCTGTTGTTGAGATGCTCTGTAAGCGATTAATCGGG 480
QY 481 CGTGGCGGGCGCTCAGTTGATGATCGGAAGAGTCACTGAGGATGCTGCAATTCGGCT 540
DB 481 CGTGGCGGGCGCTCAGTTGATGATCGGAAGAGTCACTGAGGATGCTGCAATTCGGCT 540
QY 541 GTGGAATATGTCATAGCTGAGCTCCGCTTTTATTCGAGAGCTCCGAGATTTTATAGAGGTG 600
DB 541 GTGGAATATGTCATAGCTGAGCTCCGCTTTTATTCGAGAGCTCCGAGATTTTATAGAGGTG 600

DB 541 GTGGAATATGTCATAGCTGAGCTCCGCTTTTATTCGAGAGCTCCGAGATTTTATAGAGGTG 600
QY 601 GAAGAGACACTCTCTGCTTATCATCTCCGTGGAAGCTGGTGAAGATCAGTAACCAT 660
DB 601 GAAGAGACACTCTCTGCTTATCATCTCCGTGGAAGCTGGTGAAGATCAGTAACCAT 660
QY 661 GAATTTCTTATTTGATGCGGCGGAATCAAGGCTATCTCTATTGTACAGGAAATGGCGCAG 720
DB 661 GAATTTCTTATTTGATGCGGCGGAATCAAGGCTATCTCTATTGTACAGGAAATGGCGCAG 720
QY 721 ATGCTTTCTGAGAAACGGATCATCATCTGAAGGA 753
DB 721 ATGCTTTCTGAGAAACGGATCATCATCTGAAGGA 753

RESULT 2

ACF70114

ID ACF70114 standard; DNA; 705 BP.

XX AC ACF70114;

XX AC ACF70114;

XX 20-NOV-2003 (first entry)

XX Photorhabdus luminescens nucleotide sequence #8581.

XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;

XX detection; food; gene expression; plant; animal; microorganism; toxin;

XX antibiotic; biopesticide; virulence factor; disease model; plague;

XX whooping cough; gene; ds.

XX Photorhabdus luminescens.

XX WO200294867-A2.

XX 28-NOV-2002.

XX 07-FEB-2002; 2002WO-IB003040.

XX 07-FEB-2001; 2001FR-00001659.

XX (INSP) INST PASTEUR.

XX (CNRS) CNRS CENT NAT RECH SCI.

XX Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;

XX Buchrieser C;

XX WPI; 2003-148459/14.

XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,

XX useful e.g. as therapeutic antimicrobials and agricultural pesticides.

XX Claim 2; SEQ ID NO 8581; 1205pp; French.

XX The invention relates to the isolation of genes and their encoded
XX proteins from Photorhabdus luminescens. The isolated sequences are
XX sources of probes and primers for detecting the genome of P. luminescens
XX and related species; to study polymorphisms; for gene analysis and for
XX detection/amplification of the genes. Antibodies (Ab) raised against the
XX polypeptides encoded by the genes are used for detection/identification
XX of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
XX carry a gene-containing vector are used to select compounds that
XX modulate, regulate, induce or inhibit expression of the genes in plants,
XX animals or microorganisms other than P. luminescens and are able to alter
XX response or sensitivity to toxins and antibiotics produced by P.
XX luminescens. Cells transformed to express the genes are useful for
XX recombinant production of the proteins, particularly toxins and
XX antibacterials useful as insecticides, bactericides and fungicides. The
XX genes, proteins, vectors containing the genes and Ab are also useful
XX therapeutically (to treat microbial infection by bacteria or fungi that
XX are sensitive to P. luminescens-encoded toxins or antibiotics) and as
XX biopesticides. Other uses of the genes and the proteins are as virulence
XX factors and for identifying targets of human diseases for which P.
XX luminescens is a model (particularly plague and whooping cough). This

CC sequence represents one of the isolated P. luminescens genes

XX Sequence 705 BP; 231 A; 141 C; 118 G; 215 T; 0 U; 0 Other;

Query Match 8.2%; Score 61.6; DB 10; Length 705;
Best Local Similarity 46.4%; Pred. No. 1.6e-08;
Matches 310; Conservative 0; Mismatches 349; Indels 9; Gaps 3;

```
QY 46 TTTATTCTGAGGCAATTACAGAAAACCTGCGAGGAAATATTGAAACGGCGAGGCAATGTT 105
DB 25 TTTACTGTCCAAAGTGAAACCTCTCGTTGTGACCAATATTCAAAAAGGTGATCACGCG 84
QY 106 TTGGTGGGATCAGCCCATTTAAACAGCAGGTTTTCAGAGATATATTACAGATTAAAT 165
DB 85 CTAAATAGGATAAGCCCTTTAACTCGCGTTTTCAAAAGACATATGATGAGCACTTATT 144
QY 166 GGATGGCGGAAAGCTCAAT---TAAAGGCTTTTCAGTTTACCTTTCAGGCGATGAGCG 222
DB 145 CAGTGGTCAAGTCATTAATTTCCGCAAGTCGACATATATTACCTTGTGAAGTGAAGCT 204
QY 223 GCTAATCTCTAGAGCGCTTGGAACTCCGAGAGAAAGGCTGAACGAAAGTAAGGAAA 282
DB 205 TCAGCGCTTTTAGTCGCTAGTGGAAATGATAATGTTAAAGCTATCAAAAACACATCGC 264
QY 283 GAGGTATCAGAAACAGGAGATTGCGAGAAAGCCCTTGTGGC-----TCATGGCGGG 337
DB 265 GAAATTAGAGCTCATTTACGTAACCTTGTATGTTATTTCCACAGCAACATTTGAAAAGT 324
QY 338 ATCGGAAGGCAATTCATAC-ATTTTCTGATTTTATAGATAACAAAGCTACCAAGCTGTTG 396
DB 325 AAGCAATCAGATCATCAATTTAGTGAATTTTCTCACTAAACCATGATCACTATCTCTT 384
QY 397 AGCAAGAAGTTGAACATGCAATTTTTCAGAGCGCTCAATTTTCGACATGCTTGTGTTGAC 456
DB 385 AAAACACAAGTTGAAACCGGTTTAAATGAATCAGAACTTTTAAAGGCTGCTTTGAT 444
QY 457 ATGTCCTCGTAGGCAATATCGGGCGTGGCGGGCGTCAAGTTTGTATGATGGAAGAATC 516
DB 445 ATGTCTCTCAAGCCATTAAGGCGGCTAAAGGCTACTGGGCAATATCTTGTGTCAAAT 504
QY 517 AGTGAGGATATGCTGAATTTGGCTGTGAATATGTCATAGCTAGCTGCGGCTTTTATC 576
DB 505 GACCTACAATTAGTATATAAGCGTTGCCATATATTTTCGTGAAATCTCTTTTACCTC 564
QY 577 GGAGCTCCGGATATTTTAGAGGTGGAAGAGACACTCTCTGCTTATCATCTGCTCGGTAAG 636
DB 565 AATACCCCTCGATTACTTGGGTAAAGTATTTACGTTACTTTATCACCGCCCTTGTCA 624
QY 637 CTGGGTGAGAAGATCAGTAACCATGAAATTTCTATTGTTATGCGGCGCAATCAAGGTAT 696
DB 625 ATCGAAAAGGGTTATTAAACGGTAGTTATCTCTATACAAGTAGCAGATAAACAAGTTAC 684
QY 697 CTCATTGT 704
DB 685 GGAATCGT 692
```

RESULT 3

ACF65385_2
Continuation (3 of 7) of ACF65385 from base 200001 (Photorhabdus luminescens nucleotide
WP Sequence split into 7 fragments LOCUS ACF65385 Accession ACF65385

WP	Fragment Name	Begin	End
WP	ACF65385_0	1	110000
WP	ACF65385_1	100001	210000
WP	ACF65385_2	200001	310000
WP	ACF65385_3	300001	410000
WP	ACF65385_4	400001	510000
WP	ACF65385_5	500001	610000
WP	ACF65385_6	600001	618776

Query Match 8.2%; Score 61.6; DB 10; Length 110000;
Best Local Similarity 46.4%; Pred. No. 2.1e-07;
Matches 310; Conservative 0; Mismatches 349; Indels 9; Gaps 3;

```
QY 46 TTTATTCTGAGGCAATTACAGAAAACCTGCGAGGAAATATTGAAACGGCGAGGCAATGTT 105
DB 82443 TTTACTGTCCAAAGTGAAACCTCTCGTTGTGACCAATATTCAAAAAGGTGATCACGCG 82502
QY 106 TTGGTGGGATCAGCCCATTTAAACAGCAGGTTTTCAGAGATATATTACAGATTAAAT 165
DB 82503 CTAAATAGGATAAGCCCTTTAACTCGCGTTTTCAAAAGACATATGATGAGCACTTATT 82562
QY 166 GGATGGCGGAAAGCTCAAT---TAAAGGCTTTTCAGTTTACCTTTCAGGCGATGAGCG 222
DB 82563 CAGTGGTCAAGTCATTAATTTCCGCAAGTCGACATATATTACCTTGTGAAGTGAAGCT 82622
QY 223 GCTAATCTCTAGAGCGCTTGGAACTCCGAGAGAAAGGCTGAACGAAAGTAAGGAAA 282
DB 82623 TCAGCGCTTTTAGTCGCTAGTGGAAATGATAATGTTAAAGCTATCAAAAACACATCGC 82682
QY 283 GAGGTATCAGAAACAGGAGATTGCGAGAAAGCCCTTGTGGC-----TCATGGCGGG 337
DB 82683 GAAATTAGAGCTCATTTACGTAACCTTGTATGTTATTTCCACAGCAACATTTGAAAAGT 82742
QY 338 ATCGGAAGGCAATTCATAC-ATTTTCTGATTTTATAGATAACAAAGCTACCAAGCTGTTG 396
DB 82743 AAGCAATCAGATCATCAATTTAGTGAATTTTCTCACTAAACCATGATCACTATCTCTT 82802
QY 397 AGCAAGAAGTTGAACATGCAATTTTTCAGAGCGCTCAATTTTCGACATGCTTGTGTTGAC 456
DB 82803 AAAACACAAGTTGAAACCGGTTTAAATGAATCAGAACTTTTAAAGGCTGCTTTGAT 82862
QY 457 ATGTCCTCGTAGGCAATATCGGGCGTGGCGGGCGTCAAGTTTGTATGATGGAAGAATC 516
DB 82863 ATGTCTCTCAAGCCATTAAGGCGGCTAAAGGCTACTGGGCAATATCTTGTGTCAAAT 82922
QY 517 AGTGAGGATATGCTGAATTTGGCTGTGAATATGTCATAGCTAGCTGCGGCTTTTATC 576
DB 82923 GACCTACAATTAGTATATAAGCGTTGCCATATATTTTCGTGAAATCTCTTTTACCTC 82982
QY 577 GGAGCTCCGGATATTTTAGAGGTGGAAGAGACACTCTCTGCTTATCATCTGCTCGGTAAG 636
DB 82983 AATACCCCTCGATTACTTGGGTAAAGTATTTACGTTACTTTATCACCGCCCTTGTCA 83042
QY 637 CTGGGTGAGAAGATCAGTAACCATGAAATTTCTATTGTTATGCGGCGCAATCAAGGTAT 696
DB 83043 ATCGAAAAGGGTTATTAAACGGTAGTTATCTCTATACAAGTAGCAGATAAACAAGTTAC 83102
QY 697 CTCATTGT 704
DB 83103 GGAATCGT 83110
```

RESULT 4

ACF67367_31
Continuation (32 of 57) of ACF67367 from base 3100001 (Photorhabdus luminescens nucleotide
WP Sequence split into 57 fragments LOCUS ACF67367 Accession ACF67367

WP	Fragment Name	Begin	End
WP	ACF67367_0	1	110000
WP	ACF67367_01	100001	210000
WP	ACF67367_02	200001	310000
WP	ACF67367_03	300001	410000
WP	ACF67367_04	400001	510000
WP	ACF67367_05	500001	610000
WP	ACF67367_06	600001	710000
WP	ACF67367_07	700001	810000
WP	ACF67367_08	800001	910000
WP	ACF67367_09	900001	1010000
WP	ACF67367_10	1000001	1110000
WP	ACF67367_11	1100001	1210000
WP	ACF67367_12	1200001	1310000
WP	ACF67367_13	1300001	1410000
WP	ACF67367_14	1400001	1510000
WP	ACF67367_15	1500001	1610000
WP	ACF67367_16	1600001	1710000
WP	ACF67367_17	1700001	1810000

WP	ACF67367_18	1800001	1910000	
WP	ACF67367_19	1900001	2010000	
WP	ACF67367_20	2000001	2110000	
WP	ACF67367_21	2100001	2210000	
WP	ACF67367_22	2200001	2310000	
WP	ACF67367_23	2300001	2410000	
WP	ACF67367_24	2400001	2510000	
WP	ACF67367_25	2500001	2610000	
WP	ACF67367_26	2600001	2710000	
WP	ACF67367_27	2700001	2810000	
WP	ACF67367_28	2800001	2910000	
WP	ACF67367_29	2900001	3010000	
WP	ACF67367_30	3000001	3110000	
WP	ACF67367_31	3100001	3210000	
WP	ACF67367_32	3200001	3310000	
WP	ACF67367_33	3300001	3410000	
WP	ACF67367_34	3400001	3510000	
WP	ACF67367_35	3500001	3610000	
WP	ACF67367_36	3600001	3710000	
WP	ACF67367_37	3700001	3810000	
WP	ACF67367_38	3800001	3910000	
WP	ACF67367_39	3900001	4010000	
WP	ACF67367_40	4000001	4110000	
WP	ACF67367_41	4100001	4210000	
WP	ACF67367_42	4200001	4310000	
WP	ACF67367_43	4300001	4410000	
WP	ACF67367_44	4400001	4510000	
WP	ACF67367_45	4500001	4610000	
WP	ACF67367_46	4600001	4710000	
WP	ACF67367_47	4700001	4810000	
WP	ACF67367_48	4800001	4910000	
WP	ACF67367_49	4900001	5010000	
WP	ACF67367_50	5000001	5110000	
WP	ACF67367_51	5100001	5210000	
WP	ACF67367_52	5200001	5310000	
WP	ACF67367_53	5300001	5410000	
WP	ACF67367_54	5400001	5510000	
WP	ACF67367_55	5500001	5610000	
WP	ACF67367_56	5600001	5648894	
Query Match				8.2%; Score 61.6; DB 10; Length 110000;
Best Local Similarity				46.4%; Pred. No. 2.1e-07;
Matches 310; Conservative				0; Mismatches 349; Indels 9; Gaps 3;
Qy	46	TTTATTCTGAGGCAATTAACAGAAAACCTGCAGAGAAATATTGAAACGGCGAGGCATGTT	105	
Db	87577	TTTACTGTCCAAGGTGAACCTCTCTGTGTGACCAAAATTTCAAAAAGGTGATCAGCG	87636	
Qy	106	TTGGTGGGATCAGCCCATTTAAGAGCGTTTTCAGAGGATTTATTTACAGATTAAAT	165	
Db	87637	CTAATAGGATAAGCCCTTTAATCTCGGTTTTTCAAAAGACTAAGTAGTGACCTTAT	87696	
Qy	166	GGATGGCGGAAAGCTCAATT---TAAAGCGTTTTCAGTTTACTTGAGGGCATGAGCG	222	
Db	87697	CAGTGGTCAAGTCATTATTTCCGACAGTGCACATATTTATACCTTGTGAACGTGAAGCT	87756	
Qy	223	GCTAATCTTCTAGAGCGCTTGGAACCTCCGAGAGAAAGCTGAACGAAAAGTAGAGAA	282	
Db	87757	TCAGCGCTTTTAGTCGCTAGTGGAAATGATAATGTTAAAGCTATCAAAAACACATCGC	87816	
Qy	283	GAGTATCAGAAACAGGAGATTTGACAGAAAGCCCTTTGGC---TCATGGCGGG	337	
Db	87817	GAATTAGACTCATTTACGTAACTGATTTATTTATTTCCACGACAACTGAAAAGT	87876	
Qy	338	ATCCGAAGGCAATTCATAC-ATTTTCTGATTTTATAGATAACAAAGCCTACCACTGTTG	396	
Db	87877	AAGCAATCAGATCATCAATTTAGTGACCTTTTCACTAAACCATGACTACCAATCTCT	87936	
Qy	397	AGACAAGGTTGAACATGCAATTTTTCAGAGCGCTCAATTTTCAGACTGCTTTGTTGGAC	456	
Db	87937	AAAACACAAGTTGAAAAGCGCTTTTAATGAATCAGAAATCTTTTAAAAAAGCTGCTTGAT	87996	
Qy	457	ATGTCTCGTGAAGCGAATAATCGGCGTGCAGCGGCGTTCAGTTTGTATGATGAAGAAGTC	516	

Db	87997	ATGTCCTTTCAAGCCATAAAAGGGGCGACTAAAGAGTACTGGGCAATACTTTTGTCAAAT	88056	
Qy	517	AGTGAGGATATGCTGAATTTGGCTGTGGAAATATGTATAGCTGAGCTGCCGTTTATC	576	
Db	88057	GACCTACAATTAGTATATAAAGCGTTGCCATATATTTTGCCTGAAATTCCTTTTACCTC	88116	
Qy	577	GGAGCTCCGGATATTTTAGAGGTGGAAGAGACACTCTTGTCTTATCATCGTCCTGGAAG	636	
Db	88117	AATACCCCTCGATTACTTGGGGTAAAGTATTTACGTTACTTTATCACCCCTTGTGCA	88176	
Qy	637	CTGGGTGAGAAGATCAGTAACCATGAATTTTCTATTTGTATGCGCGCAATCAAGGTTAT	696	
Db	88177	ATCGGAAGAGGTTATTTAACGGTAGTTATCTCTATACAGTACAGATAACAAAGTTAC	88236	
Qy	697	CTCATTTGT 704		
Db	88237	GGAATCGT 88244		
RESULT 5				
ACF66408				
ID ACF66408 standard; DNA; 651 BP.				
XX	AC	ACF66408;		
XX	AC	ACF66408;		
DT	20-NOV-2003 (first entry)			
XX	Photorhabdus luminescens nucleotide sequence #4875.			
XX	Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;			
KW	detection; food; gene expression; plant; animal; microorganism; toxin;			
KW	antibiotic; biopesticide; virulence factor; disease model; plague;			
KW	whooping cough; gene; ds.			
XX	Photorhabdus luminescens.			
OS	WO200294867-A2.			
PN	28-NOV-2002.			
PD	07-FEB-2002; 2002WO-IB003040.			
XX	07-FEB-2001; 2001FR-00001659.			
XX	(INSP) INST PASTEUR.			
PA	(CNRS) CNRS CENT NAT RECH SCI.			
XX	Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;			
PI	Buchrieser C;			
XX	WPI; 2003-148459/14.			
DR	Claim 28; SEQ ID NO 4875; 1205pp; French.			
XX	The invention relates to the isolation of genes and their encoded			
CC	proteins from Photorhabdus luminescens. The isolated sequences are			
CC	sources of probes and primers for detecting the genome of P. luminescens			
CC	and related species; to study polymorphisms; for gene analysis and for			
CC	detection/amplification of the genes. Antibodies (Ab) raised against the			
CC	polypeptides encoded by the genes are used for detection/identification			
CC	of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that			
CC	carry a gene-containing vector are used to select compounds that			
CC	modulate, regulate, induce or inhibit expression of the genes in plants,			
CC	animals or microorganisms other than P. luminescens and are able to alter			
CC	response or sensitivity to toxins and antibiotics produced by P.			
CC	luminescens. Cells transformed to express the genes are useful for			
CC	recombinant production of the proteins, particularly toxins and			
CC	antibacterials useful as insecticides, bactericides and fungicides. The			
CC	genes, proteins, vectors containing the genes and Ab are also useful			

CC	therapeutically (to treat microbial infection by bacteria or fungi that
CC	are sensitive to <i>P. luminescens</i> -encoded toxins or antibiotics) and as
CC	biopesticides. Other uses of the genes and the proteins are as virulence
CC	factors and for identifying targets of human diseases for which <i>P.</i>
CC	<i>luminescens</i> is a model (particularly plague and whooping cough). This
CC	sequence represents one of the isolated <i>P. luminescens</i> genes
XX	
SQ	Sequence 651 BP; 229 A; 126 C; 97 G; 198 T; 0 U; 1 Other;
	Query Match 7.8%; Score 58.4; DB 10; Length 651;
	Best Local Similarity 48.0%; Pred. No. 1.6e-07;
	Matches 167; Conservative 0; Mismatches 181; Indels 0; Gaps 0;
Qy	357 ATTTTCGATTTTATAGATAACAAGCCTACCACTGTTGAGACAAGATTGAACATGC 416
Db	140 ATTTAGTGACTTTTTCACATAAACCACTGACTACCAATCTCTTAAAAACACAAGTTGAAAACGC 199
Qy	417 ATTTTTTGAGCAGCCTCATTTTTCACATGCTTGTGTCGACATGCTCTCGTGAAGCGATAAT 476
Db	200 GTTTTAATGAATCAGAAATCTTTTAAAAAAGCTGCTCTTGATATGCTCTTCAAGCCATAAA 259
Qy	477 CGGCGCTGCGGGCGCTCAGTTTGATGATGGAAGAACTCAGTGAGGATATCGTGAATTT 536
Db	260 AGGCGGACTAAAGGTACTGGCAATCTTTGGTCAAATTCACCTACAATTAGTATATAA 319
Qy	537 GGCTGTGGAATATGCTATAGCTGAGCTGCCGCTTTTATCGGAGCTCCGGATATTTTGA 596
Db	320 AGCGTGGCCATATATTTTCGCTGAATTCCTTTTACCTCAATACCCCTCGATTACTTGG 379
Qy	597 GGTGGAAGAGACATCCTTGTCTATCATCTGTCGCTGGAAGCTGGTGAGAGATCAGTAA 656
Db	380 GGTAAAGTATTTCTACGTACTTTTATCACCGCCCTTGGTCAATCGAAGAAAGGTTATTAA 439
Qy	657 CCATGAATTTTCTATTTGTATGCGCGCAATCAAGGGGTATCTCATGT 704
Db	440 CGGTAGTTATCCYATACAGTAGCAGATAACAAGTTACGGAATCGT 487
RESULT 6	
ADA71938	ID ADA71938 standard; DNA; 2000 BP.
XX	ADA71938;
AC	
XX	20-NOV-2003 (first entry)
DT	
XX	
XX	Rice gene, SEQ ID 5263.
DE	
XX	Plant; bacterial infection; fungal infection; viral infection; rice;
KW	gene; ds.
KW	
XX	
OS	Oryza sativa.
XX	
FN	WO2003000898-A1.
XX	
PD	03-JAN-2003.
XX	
XX	22-JUN-2001; 2001WO-IB001105.
PF	
XX	
PR	22-JUN-2001; 2001WO-IB001105.
XX	
XX	(SYGN) SYNGENTA PARTICIPATIONS AG.
PA	
XX	
PI	Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI	Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX	
XX	WPI; 2003-175290/17.
DR	
XX	
PT	Identifying at least one gene involved in plant resistance or response to
PT	pathogenic infection for conferring resistance or tolerance to a plant to
PT	bacterial, fungal or viral infection by determining or detecting plant
PT	gene expression.
XX	

RESULT 9

AA42063_15
Continuation (16 of 19) of AA42063 from base 1500001 (Haemophilus influenzae complete g.
WP Sequence split into 19 fragments LOCUS AA42063 Accession Aat42063
Fragment Name Begin End
WP AA42063_00 1 110000
WP AA42063_01 100001 210000
WP AA42063_02 200001 310000
WP AA42063_03 300001 410000
WP AA42063_04 400001 510000
WP AA42063_05 500001 610000
WP AA42063_06 600001 710000
WP AA42063_07 700001 810000
WP AA42063_08 800001 910000
WP AA42063_09 900001 1010000
WP AA42063_10 1000001 1110000
WP AA42063_11 1100001 1210000
WP AA42063_12 1200001 1310000
WP AA42063_13 1300001 1410000
WP AA42063_14 1400001 1510000
WP AA42063_15 1500001 1610000
WP AA42063_16 1600001 1710000
WP AA42063_17 1700001 1810000
WP AA42063_18 1800001 1930121

Query Match 5.0%; Score 38; DB 2; Length 110000;

Best Local Similarity 55.2%; Pred. No. 5.6;

Matches 74; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 244 GGAATCCGAGAGGCTGACGAAAGTAGGAAAGGATATCAGCAACAGGAGA 303
DB 19210 GAATTTCAAAAATGATTCGCGAAGAACCAACGTCAAATTCGAAACGGGAAGTCGT 19269
QY 304 TTGTCAGAAAGAGCCCTTGCGCTCATGCGGGGATCCGAGCGATTTCATACATTTTCT 363
DB 19270 TTACCTGATACATCATTCGCTGCTGTGGCGGTGCTCGAATGCAATGCTATGTTTCT 19329
QY 364 GATTTTATAGATAA 377
DB 19330 GATTTTATGATGA 19343

RESULT 10

ADH10017_0
WP Sequence split into 4 fragments LOCUS ADH10017 Accession Adh10017

Fragment Name Begin End
WP ADH10017_0 1 110000
WP ADH10017_1 100001 210000
WP ADH10017_2 200001 310000
WP ADH10017_3 300001 365186
ID ADH10017 standard; DNA; 365186 BP.

AC ADH10017;

XX XX

DT 11-MAR-2004 (first entry)

XX Human chromosome 2p21-22 fragment containing the DRIP gene.

XX human; chromosome 2; hyperplasia; tumour; thyroid; cancer; break point;

KW chromosomal band 2p21-22; DRIP; cytostatic; thymimetic;

KW chromosome 7p15; ds.

XX XX

XX Synthetic.

OS Homo sapiens.

XX XX

FH Key Location/Qualifiers

FT exon 1. .108

FT FT /*tag= a

FT FT /number= 1

FT FT 3623. .3722

FT FT /*tag= b

FT FT /number= 2

FT FT 3993. .4087

FT FT exon

FT FT /*tag= c
FT FT /number= 3
FT FT 5085. .5215
FT FT /*tag= d
FT FT /number= 4
FT FT 9037. .9185
FT FT /*tag= e
FT FT /number= 5
FT FT 9305. .9337
FT FT /*tag= f
FT FT /number= 6
FT FT 9590. .9638
FT FT /*tag= g
FT FT /number= 7
FT FT 14110. .14297
FT FT /*tag= h
FT FT /number= 8
FT FT 17431. .17525
FT FT /*tag= i
FT FT /number= 9
FT FT 18796. .19016
FT FT /*tag= j
FT FT /number= 10
FT FT 21011. .21702
FT FT /*tag= k
FT FT /number= 11
FT FT 23046. .23224
FT FT /*tag= l
FT FT /number= 12
FT FT 24176. .24331
FT FT /*tag= m
FT FT /number= 13
FT FT 25528. .25650
FT FT /*tag= n
FT FT /number= 14
FT FT 29217. .29340
FT FT /*tag= o
FT FT /number= 15
FT FT 35653. .35804
FT FT /*tag= p
FT FT /number= 16
FT FT 39483. .39693
FT FT /*tag= q
FT FT /number= 17
FT FT 43699. .43834
FT FT /*tag= r
FT FT /number= 18
FT FT 44113. .44249
FT FT /*tag= s
FT FT /number= 19
FT FT 46670. .46828
FT FT /*tag= t
FT FT /number= 20
FT FT 54722. .54879
FT FT /*tag= u
FT FT /number= 21
FT FT 68050. .68159
FT FT /*tag= v
FT FT /number= 22
FT FT 87258. .87390
FT FT /*tag= w
FT FT /number= 23
FT FT 90303. .90416
FT FT /*tag= x
FT FT /number= 24
FT FT 97083. .97205
FT FT /*tag= y
FT FT /number= 25
FT FT 110713. .110804
FT FT /*tag= z
FT FT /number= 26
FT FT 165736. .165825
FT FT /*tag= aa

FT exon /number= 27
FT 167807..167938
FT /*tag= ab
FT /number= 28
FT 197899..198067
FT /*tag= ac
FT /number= 29
FT 251801..251916
FT /*tag= ad
FT /number= 30
FT 275498..275592
FT /*tag= ae
FT /number= 31
FT 302825..303204
FT /*tag= af
FT /number= 32
FT 303816..303934
FT /*tag= ag
FT /number= 33
FT 304270..304342
FT /*tag= ah
FT /number= 34
FT 308977..309130
FT /*tag= ai
FT /number= 35
FT 316142..316273
FT /*tag= aj
FT /number= 36
FT 363156..363325
FT /*tag= ak
FT /number= 37
FT 364695..365186
FT /*tag= al
FT /number= 38
XX WO2003093310-A1.
XX
XX
XX
XX 13-NOV-2003.
XX
XX 02-MAY-2003; 2003WO-EP004642.
XX
XX 01-MAY-2002; 2002DE-01019413.
XX 14-SEP-2002; 2002DE-01042705.
XX
XX (UYBR-) UNIV BREMEN.
XX
XX Bullerdiel J;
XX
XX WPI; 2003-854480/79.
XX
XX New nucleic acid from human chromosome 2, useful for treatment and
XX diagnosis of thyroid disease, especially cancer, also related
XX polypeptides and modulators.
XX
XX Claim 1; SEQ ID NO 12; 461pp; German.
XX
XX This invention describes a novel nucleic acid from human chromosome 2
XX that shows altered expression in hyperplasia and/or tumours, especially
XX of the thyroid. The invention describes a method for preparing nucleic
XX acid that can be detected in thyroid cancers where these contain an
XX aberration of the break point in chromosomal band 2p21-22. The invention
XX also describes constructs comprising a fragment of the human DRIP gene
XX and its splice variants, one of which contains all of exons 1-38 while
XX the other lacks exons 27 and 28. Some tumours contain a fusion protein of
XX DRIP, including a small segment from chromosome 3. The products of the
XX invention have cytostatic and thyromimetic activity and are used for the
XX inhibition of nucleic acid expression by antisense, ribozyme or RNA
XX interference (RNAi) methods. This sequence represents the human
XX chromosome 2p21-22 DNA containing the region encoding the DRIP protein
XX encoded by exons 1-38, described in the disclosure of the invention.
XX
XX Sequence 365186 BP; 103587 A; 71408 C; 76380 G; 113811 T; 0 U; 0 Other;

Query Match 4.9%; Score 36.8; DB 10; Length 110000;
Best Local Similarity 54.4%; Pred. No. 13;
Matches 74; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
QY 328 CATGGCGGGATCCGAAGGCGATTTCATACATTTTCTGATTTTATAGATAACAAGCCTAC 387
DB 107307 CATAGAGTTTACTCTGTGACCAAGCAATACTCTTCTAGGTATATAGTCATGAGAAATGAA 107366
QY 388 CAGCTGTTGAGACAGAAGTTGAACATGCTATTTTGGACAGCCTCATTTTCGACATGCT 447
DB 107367 AACATGTTTACACAAAACCTTGATCATGAATATTTCTAGCAAAATATTTCATAATATTC 107426
QY 448 TGTTCGACATGCTCTC 463
DB 107427 AAATTGGAATAGCCC 107442
RESULT 11
ADH10017.1
Continuation (2 of 4) of ADH10017 from base 100001 (Human chromosome 2p21-22 fragment con
WP Sequence split into 4 fragments LOCUS ADH10017 Accession Adh10017
WP Fragment Name Begin End
WP ADH10017_0 1 110000
WP ADH10017_1 100001 210000
WP ADH10017_2 200001 310000
WP ADH10017_3 300001 365186
Query Match 4.9%; Score 36.8; DB 10; Length 110000;
Best Local Similarity 54.4%; Pred. No. 13;
Matches 74; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
QY 328 CATGGCGGGATCCGAAGGCGATTTCATACATTTTCTGATTTTATAGATAACAAGCCTAC 387
DB 7307 CATAGAGTTTACTCTGTGACCAAGCAATACTCTTCTAGGTATATAGTCATGAGAAATGAA 7366
QY 388 CAGCTGTTGAGACAGAAGTTGAACATGCTATTTTGGACAGCCTCATTTTCGACATGCT 447
DB 7367 AACATGTTTACACAAAACCTTGATCATGAATATTTCTAGCAAAATATTTCATAATATTC 7426
QY 448 TGTTCGACATGCTCTC 463
DB 7427 AAATTGGAATAGCCC 7442
RESULT 12
AAQ75343
ID AAQ75343 standard; cDNA to mRNA; 3048 BP.
XX
XX AAQ75343;
XX
XX 16-OCT-1995 (first entry)
XX
XX Soybean phosphoenolpyruvate carboxylase cDNA to mRNA.
XX
XX Soybean; phosphoenolpyruvate carboxylase; protein production;
XX fat production; transgenic crop plants; ds.
XX
XX Glycine max.
XX
XX Key Location/Qualifiers
XX CDS 28..2931
XX /*tag= a
XX
XX JF06319567-A.
XX
XX 22-NOV-1994.
XX
XX 27-SEP-1991; 91JP-00274950.
XX
XX 27-SEP-1991; 91JP-00274950.
XX
XX (NORQ) NORINSUISANSO SHOKUHIN SOGO.
XX PA (MITS-) MITSUI GYOSAI SHOKUBUTSU BIO KENKYUSHO.
XX PA

XX WPI; 1995-040320/06.
DR P-PSDB; AAR67428.
XX
XX Phospho-enol pyruvate carboxylase gene derived soy beans - to provide
PT increased protein or fat production in transgenic crop plants.
XX
XX Claim 1; Page 5-11; 19pp; Japanese.
XX
XX AAQ75343 encodes AAR67428 soybean phosphoenolpyruvate carboxylase (PEPC).
CC Expression of PEPC increases protein prodn., whilst inhibition of PEPC
CC expression increases fat prodn. Therefore PEPC can be used to modulate
CC protein and fat prodn. in transgenic crop plants
XX
XX Sequence 3048 BP; 876 A; 605 C; 724 G; 843 T; 0 U; 0 Other;
SQ
Query Match 4.8%; Score 36.4; DB 2; Length 3048;
Best Local Similarity 51.9%; Pred. No. 2.9;
Matches 82; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
QY 502 ATGATGGAAGAAGTCAGTGGAGGATATGCTGGAATTTGGCTGTGGAATATGTCATAGCTGAG 561
Db 2446 ATGCTGCAAGAGATGTACAATCAATGCGCTTCTTTAGGCTCACACTTGATTTGGTGAA 2505
QY 562 CTGCGGTTTTATCGAGCTCGGATATTTAGAGTGGGAAGACACACTCTTCTGCTTAT 621
Db 2506 ATGGTGTGTTGCCAAGAGATCGGAAATTTGCCGCTCTGAATGATAGACTCTTGTTC 2565
QY 622 CATCGTCCGTGGAGCTGGGTGAGAAGATCAGTAACCA 659
Db 2566 AAGGATCTGTGCGGTTGGGGATCAATTGAGGAACAA 2603
RESULT 13
AAQ75344
ID AAQ75344 standard; cDNA to mRNA; 3211 BP.
XX
AC AAQ75344;
XX
XX 16-OCT-1995 (first entry)
DT
XX
DE Soybean phosphoenolpyruvate carboxylase cDNA to mRNA.
XX
KW Soybean; phosphoenolpyruvate carboxylase; protein production;
KW fat production; transgenic crop plants; ds.
XX
XX Glycine max.
XX
FH Key Location/Qualifiers
FT CDS 117..3020
FT /*tag= a
XX
XX JP06319567-A.
XX
XX 22-NOV-1994.
XX
XX 27-SEP-1991; 91JP-00274950.
XX
XX 27-SEP-1991; 91JP-00274950.
XX
XX (NORQ) NORINSUISANSHO SHOKUHN SOGO.
XX (MITS-) MITSUI GYOSAI SHOKUBUTSU BIO KENKYUSHO.
XX
XX WPI; 1995-040320/06.
DR P-PSDB; AAR67429.
XX
XX Phospho-enol pyruvate carboxylase gene derived soy beans - to provide
PT increased protein or fat production in transgenic crop plants.
XX
XX Claim 1; Page 13-19; 19pp; Japanese.
XX
XX AAQ75344 encodes AAR67429 soybean phosphoenolpyruvate carboxylase (PEPC).
CC Expression of PEPC increases protein prodn., whilst inhibition of PEPC

CC expression increases fat prodn. Therefore PEPC can be used to modulate
CC protein and fat prodn. in transgenic crop plants
XX
XX Sequence 3211 BP; 910 A; 651 C; 747 G; 903 T; 0 U; 0 Other;
SQ
Query Match 4.8%; Score 36.4; DB 2; Length 3211;
Best Local Similarity 51.9%; Pred. No. 3;
Matches 82; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
QY 502 ATGATGGAAGAAGTCAGTGGAGGATATGCTGGAATTTGGCTGTGGAATATGTCATAGCTGAG 561
Db 2535 ATGCTGCAAGAGATGTACAATCAATGCGCTTCTTTAGGCTCACACTTGATTTGGTGAA 2594
QY 562 CTGCGGTTTTATCGAGCTCGGATATTTAGAGTGGGAAGACACACTCTTCTGCTTAT 621
Db 2595 ATGGTGTGTTGCCAAGAGATCGGAAATTTGCCGCTCTGAATGATAGACTCTTGTTC 2654
QY 622 CATCGTCCGTGGAGCTGGGTGAGAAGATCAGTAACCA 659
Db 2655 AAGGATCTGTGCGGTTGGGGATCAATTGAGGAACAA 2692
RESULT 14
ABA03041_24/c
Continuation (25 of 30) of ABA03041 from base 2400001 (Listeria monocytogenes EGD-e genom
WP Sequence split into 30 fragments LOCUS ABA03041 Accession ABA03041
Fragment Name Begin End
WP ABA03041_00 1 110000
WP ABA03041_01 100001 210000
WP ABA03041_02 200001 310000
WP ABA03041_03 300001 410000
WP ABA03041_04 400001 510000
WP ABA03041_05 500001 610000
WP ABA03041_06 600001 710000
WP ABA03041_07 700001 810000
WP ABA03041_08 800001 910000
WP ABA03041_09 900001 1010000
WP ABA03041_10 1000001 1110000
WP ABA03041_11 1100001 1210000
WP ABA03041_12 1200001 1310000
WP ABA03041_13 1300001 1410000
WP ABA03041_14 1400001 1510000
WP ABA03041_15 1500001 1610000
WP ABA03041_16 1600001 1710000
WP ABA03041_17 1700001 1810000
WP ABA03041_18 1800001 1910000
WP ABA03041_19 1900001 2010000
WP ABA03041_20 200001 2110000
WP ABA03041_21 2100001 2210000
WP ABA03041_22 2200001 2310000
WP ABA03041_23 2300001 2410000
WP ABA03041_24 2400001 2510000
WP ABA03041_25 2500001 2610000
WP ABA03041_26 2600001 2710000
WP ABA03041_27 2700001 2810000
WP ABA03041_28 2800001 2910000
WP ABA03041_29 2900001 2944528
Query Match 4.8%; Score 36.2; DB 6; Length 110000;
Best Local Similarity 45.8%; Pred. No. 21;
Matches 128; Conservative 0; Mismatches 153; Indels 0; Gaps 0;
QY 101 ATGTTTTGGTGGGATCAGCCCATTTAAACAGCAGGTTTTTACAGAGGATTTATATTACAGAT 160
Db 25556 AGGATTTGGAGGCGGGGATTAATGGAGGAAGTTTACTACGGAATTAATGTTTACAT 25497
QY 161 TAATTGGATGGCGAAAGCTCAATTTAAAGCGTTTCAGTTTTTACTTTGCGAGGCGATGAGG 220
Db 25496 TAACTAATAATCCAAAGTTTCAAGAAGTAGCAGAATTTGGTATATTGGAACGAGGAACAA 25437
QY 221 CGGCTAATCTTCTAGAGCGCTTGGAACTCCGAGAGGAAGGCTGACGAAAGTAAAGTAAGGA 280
Db 25436 TTGTTCAATGTCAGAAACAAATGAGTATGTGATGCAATTGACGAGGAAGTGTAAACGC 25377

QY 281 AAGAGGTATCAGCAACAGGAGATTTCAGAAAGAGCCCTTGGCTCATGGCGGGATC 340
|||
Db 25376 AATGTGCTGCAAGAAACAGTCTTGGGAAAAGGAAGCATCTCGGTTTACTCGAGAAG 25317
|||
QY 341 CGAAGCGGATTATACATTTCTCGATTTTATAGATAACAAA 381
|||
Db 25316 AGACAAAATCCAACCCCTTCTAAACGGTAGCCTGGAAA 25276
|||

RESULT 15
ACF72025
ID ACF72025 standard; DNA: 1338 BP.
XX
AC ACF72025;
XX
DT 20-NOV-2003 (first entry)
XX
DE Photorhabdus luminescens nucleotide sequence #10492.
XX
KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW detection; food; gene expression; plant; animal; microorganism; toxin;
KW antibiotic; biopesticide; virulence factor; disease model; plague;
KW whooping cough; gene; ds.
XX
OS Photorhabdus luminescens.
XX
PN WO200294867-A2.
XX
PD 28-NOV-2002.
XX
PF 07-FEB-2002; 2002WO-1B003040.
XX
PR 07-FEB-2001; 2001FR-00001659.
XX
PA (INSP) INST PASTEUR.
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
XX Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
PI Buchrieser C;
XX
XX WPI; 2003-148459/14.
XX
XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX
PS Claim 2; SEQ ID NO 10492; 1205pp; French.
XX
XX The invention relates to the isolation of genes and their encoded
CC proteins from Photorhabdus luminescens. The isolated sequences are
CC sources of probes and primers for detecting the genome of P. luminescens
CC and related species; to study polymorphisms; for gene analysis and for
CC detection/amplification of the genes. Antibodies (Ab) raised against the
CC polypeptides encoded by the genes are used for detection/identification
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC carry a gene-containing vector are used to select compounds that
CC modulate, regulate, induce or inhibit expression of the genes in plants,
CC animals or microorganisms other than P. luminescens and are able to alter
CC response or sensitivity to toxins and antibiotics produced by P.
CC luminescens. Cells transformed to express the genes are useful for
CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luminescens genes
XX

SQ Sequence 1338 BP; 344 A; 269 C; 367 G; 358 T; 0 U; 0 Other;
Query Match 4.8%; Score 36; DB 10; Length 1338;

Best Local Similarity 51.9%; Pred. No. 2.6;
Matches 81; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
QY 392 TGTGAGACAAGAAGTTGAACATGCATTTTTCAGCAGCGCTCATTTTCGACATGCTTGT 451
|||
Db 590 TCGCGGAACAACCTCCTTGAGTCAGAACTGTTTGGTCATGCTAAAGGTGCATTTTACCGGTG 649
|||
QY 452 TGGACATGTCGTGAAGCGATAATCGGGCGTSCGCGGGCGTCAGTTTTCATGATGGAAG 511
|||
Db 650 CGGTCAGCAGCCGTGAAGGTTATTCTTGTGCGCAGGGGGGAACGCTGTTTCTTGATG 709
|||
QY 512 AAGTCAGTGAGGATATGCTGAATTTTGGCTGTGGAAT 547
|||
Db 710 AAATCGTGTATATGCCACAGGCTTTTCAGGTTAAAT 745
|||

Search completed: December 10, 2004, 13:46:26
Job time : 424.058 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2004, 13:19:06 ; Search time 76.1418 Seconds
(without alignments)
7029.309 Million cell updates/sec

Title: US-10-627-124-7
Perfect score: 753
Sequence: 1 gtaagagagatgccggaat.....aacggatcacatctgaagga 753

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match: 0%
Maximum Match: 100%
Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	48.6	6.5	7218	1	US-08-232-463-14
2	41.2	5.5	1659	1	US-08-231-729B-1
3	41.2	5.5	1659	1	US-08-231-729B-2
4	38	5.0	1830121	4	US-09-557-884-1
5	38	5.0	1830121	4	US-09-643-990A-1
6	38	5.0	1830121	4	US-10-329-960-1
C 7	36.2	4.8	399	4	US-09-621-976-8976
C 8	34.8	4.6	640681	4	US-09-780-988-1
9	34	4.5	395	4	US-09-894-844-45
C 10	33.8	4.5	1230025	4	US-09-198-452A-1
11	33.8	4.5	1664976	4	US-08-916-421B-1
12	33.8	4.5	1664976	4	US-09-692-570-1
C 13	33.4	4.4	879	4	US-09-489-039A-3818
C 14	33.2	4.4	337	4	US-09-270-767-2403
C 15	33.2	4.4	337	4	US-09-270-767-17685
16	32.6	4.3	1500	4	US-09-248-796A-545
17	32.4	4.3	234	4	US-09-489-039A-6410
18	32.4	4.3	1467	4	US-09-710-279-2307
19	32.4	4.3	1485	3	US-09-134-001C-2187
C 20	32.4	4.3	2291	4	US-09-023-655-627
21	32.4	4.3	2980	4	US-09-710-279-4242
22	32.4	4.3	3840	4	US-09-710-279-4013
23	32	4.2	1141	4	US-09-806-708B-22
C 24	31.8	4.2	1882	1	US-08-431-080-15
25	31.8	4.2	1882	1	US-08-431-080-29
C 26	31.8	4.2	1882	2	US-08-938-534-15
27	31.8	4.2	1882	2	US-08-938-534-29

C	28	31.8	4.2	1882	3	US-09-345-294-15	Sequence 15, Appl
	29	31.8	4.2	1882	3	US-09-345-294-29	Sequence 29, Appl
C	30	31.8	4.2	2251	4	US-09-549-872B-16	Sequence 16, Appl
C	31	31.8	4.2	3180	4	US-09-549-872B-6	Sequence 6, Appl
C	32	31.8	4.2	11207	4	US-09-549-872B-2	Sequence 2, Appl
33	31.8	4.2	1230025	4	US-09-198-452A-1	Sequence 1, Appl	
34	31.6	4.2	679	4	US-09-634-238-78	Sequence 78, Appl	
35	31.6	4.2	832	4	US-09-621-976-2813	Sequence 2813, Ap	
36	31.6	4.2	2901	4	US-09-606-312-1	Sequence 1, Appl	
37	31.6	4.2	2982	4	US-09-710-279-4062	Sequence 4062, Ap	
38	31.6	4.2	4235	4	US-09-710-279-4041	Sequence 4041, Ap	
C	39	31.4	4.2	474	4	US-09-621-976-18033	Sequence 18033, A
40	31.4	4.2	495	4	US-09-710-279-809	Sequence 809, App	
41	31.4	4.2	435	4	US-09-710-279-2129	Sequence 2129, Ap	
42	31.4	4.2	1101	3	US-09-134-001C-842	Sequence 842, App	
C	43	31.4	4.2	1141	4	US-09-806-708B-22	Sequence 22, Appl
C	44	31.4	4.2	2463	4	US-09-489-039A-2971	Sequence 2971, Ap
45	31.4	4.2	3244	4	US-09-710-279-3567	Sequence 3567, Ap	

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: ptzgt-File
US-08-232-463-14
Query Match 6.5%; Score 48.6; DB 1; Length 7218;

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1659 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1659
OTHER INFORMATION: for codons 28, 32, 112, 130, 142, 190, 212,
217, 222, 266, 329, 336, 396, 436, 512, and 532, if the 3'
OTHER INFORMATION: nucleotide is T or C, then the 5' nucleotide is C; and if the
OTHER INFORMATION: nucleotide is A, then the 3' nucleotide is A or G
US-08-231-729B-2

Query Match 5.5%; Score 41.2; DB 1; Length 1659;
Best Local Similarity 30.08; Pred. No. 0.0062;
Matches 93; Conservative 39; Mismatches 178; Indels 0; Gaps 0;

QY 304 TTTGAGAAAGAGCCCTTGGCTCATGCGGGGATCCGAGGCGATTTCATACATTTTCT 363
DB 1255 TCNGNGAYATHGNTAYTTVGAYGARGAYGNCAYGNTAYATHGNTGAYMGNYTNAAR 1314
QY 364 GATTTTATAGATACAAAGCCTACCAGCTGTGAGACAAGAGTTGAACATGCAATTTTT 423
DB 1315 TCNYTNAATHAARTAYAAAGGNTAYCARGTNCNCNGCNGARYTNGARGCNYTNYTNTN 1374
QY 424 GAGCAGCCTCATTTTCACATGCTTGTGACATGCTCGTGAAGCGATAATCGGGCGT 483
DB 1375 CARCAYCCTTATGARGAYGNGGNGTNCNGNGTNCNGAYGARGTNGCNGNGAY 1434
QY 484 GCGGGGCGCTCAGTTTGATGATGGAAGAGTCACTGAGGATATGCTGAATTTGGCTGTG 543
DB 1435 YTNCCNGGNGGNGTNGTNYTNAARGARGNAARTCNATHACNGARAARGARATHCAR 1494
QY 544 GAATATGCTAGCTAGCTGCGCTTTTATCGGAGCTCGGATATTTAGAGTGGAA 603
DB 1495 GAYTAYTNGGNGCARGTNCACNTCNCNAAARYTNGGNGGNGTNGARTTYGTN 1554
QY 604 GAGACACTCC 613
DB 1555 AARGAGTNC 1564

RESULT 4
US-09-557-884-1
Sequence 1, Application US/09557884
Patent No. 6506581
GENERAL INFORMATION:
APPLICANT: Fleischmann et al.
TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/557,884
FILING DATE: 25-Apr-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB186P1C1

FILING DATE: JUN-5-1995
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB186P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Query Match 5.0%; Score 38; DB 4; Length 1830121;
Best Local Similarity 55.2%; Pred. No. 2.9;
Matches 74; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 244 GGAACCTCCGAGAGAAAGGCTGAACGAAAGTAAGGAAAGAGGTATCACGAAACAGGAGA 303
DB 1519210 GAATTTCAAAATGATTGGCGAAGAACCAACGTCAAATTCGAAACGGAAGGTCGT 1519269
QY 304 TTTGCGAAGAGCCCTTGTGGCTCATGCGGGGATCCGAGGCGATTTCATACATTTTCT 363
DB 1519270 TTACCTGATACAGTCATTGCTGCTGTTGGCGGTGCTCGAATGCAATGATGTTGCT 1519329
QY 364 GATTTATAGATAA 377
DB 1519330 GATTTTATGATGA 1519343

RESULT 5
US-09-643-990A-1
Sequence 1, Application US/09643990A
Patent No. 6528289
GENERAL INFORMATION:
APPLICANT: Robert D. Fleischmann
Owen White
Hamilton O. Smith
J. Craig Venter
TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville,
STATE: MD
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,990A
FILING DATE: 23-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION NUMBER: 08/487,429
FILING DATE: 1995-06-07
APPLICATION NUMBER: 08/426,787
FILING DATE: 1995-04-21
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB186P1C1

```
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-610-5790
; TELEFAX: 310-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match          5.0%; Score 38; DB 4; Length 1830121;
Best Local Similarity 55.2%; Pred. No. 2.9;
.Matches 74; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 244 GGAACGCCGAGAGGCTGCAACAAAGTAAGGAAGAGGATATCAGCAACAGGAGA 303
Db 1519210 GAATTTCAAAATGATTGGCGAAGAACCAACGTCAAATTCGACGGAAGTCGT 1519269

QY 304 TTGCGAAGAGCCCTTGTGGCTCATGGCGGGATCCGAGGCGATTATACATTTTCT 363
Db 1519270 TTACCTGATACAGTCATGCTGCTGTTGGCGTGGCTCGAATTCGAATTTGCT 1519329

QY 364 GATTTTATAGATA 377
Db 1519330 GATTTTATTATGCA 1519343

RESULT 6
US-10-329-960-1
; Sequence 1, Application US/10329960
; Patent No. 6742927
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag
; Patent No. 6742927
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB186P1
; CURRENT APPLICATION NUMBER: US/10/329,960
; CURRENT FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (40808)..(40810)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44416)..(44416)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44905)..(44905)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45593)..(45593)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45732)..(45732)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (47036)..(47036)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51334)..(51334)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51602)..(51602)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51786)..(51786)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51805)..(51805)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (55369)..(55369)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65309)..(65309)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65313)..(65313)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (80024)..(80024)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (100091)..(100091)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
```

```
; NAME/KEY: misc_feature
; LOCATION: (102696)..(102696)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (105121)..(105121)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (107248)..(107248)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (117136)..(117136)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (119750)..(119750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (119924)..(119924)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (120038)..(120038)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (121344)..(121344)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (122167)..(122167)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (122336)..(122336)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (131340)..(131340)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (131360)..(131360)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (139910)..(139910)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (140398)..(140398)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (142750)..(142750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145058)..(145058)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145171)..(145171)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145942)..(145942)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
```

```
; LOCATION: (147197)..(147197)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (150841)..(150841)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (152500)..(152500)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature

Query Match      5.0%; Score 38; DB 4; Length 1830121;
Best Local Similarity 55.2%; Pred. No. 2.9;
Matches 74; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 244 GGAACCTCCGAGAGGAGCTGAACAAAGTAAGGAAGAGGTATCAGCAACAGAGA 303
DB 1519210 GAATTTCAAAAAATGATTGGCGAAGAAACCAACGTCAAATTCGAAACGGGAAGTCGT 1519269
QY 304 TTTGCAGAAAGAGCCCTTGTGGCTCATGGCGGGATCCGAAGCGGATTTCATACATTTC 363
DB 1519270 TTACCTGATACAGTCATTGCTGCTGTGGCGGTGCGATCGAATTCGATGTTGCT 1519329
QY 364 GATTTTATAGATAA 377
DB 1519330 GATTTTATTGATGA 1519343

RESULT 7
US-09-621-976-8976/c
; Sequence 8976, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 8976
; LENGTH: 399
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-621-976-8976

Query Match      4.8%; Score 36.2; DB 4; Length 399;
Best Local Similarity 10.0%; Pred. No. 0.12;
Matches 35; Conservative 158; Mismatches 156; Indels 0; Gaps 0;

QY 357 ATTTCTGATTTTATAGATAACAAAGCTACAGCTGTGGAGACAAGTTGAACATGC 416
DB 356 AWCKWWSGAMKWRGTGKWKYVYMTSRGYSRYKTSAMMGRAMKRRKTKMYMKKG 297
QY 417 ATTTTGTGACAGCCTCATTTTCGACATGCTTTTGGACATGTCGTCGAAGCGATAAT 476
DB 296 GKXGSGTYMAMRSRRGSTGRWSYRAMWRGSKWGGSYRMAGYSSRWSRWSYAMWRK 237
QY 477 CGGGCGTGGCGGGCGCTCAGTTTGATGAGGAAGTCAAGTGAGGATATGCTGAATTT 536
DB 236 KMTCKWGRSSWSRSTGYAMMYKSWCTSRKMYKKRKKWRKCTSKTCTYRGST 177
QY 537 GGCTGTGGAATATGTCATAGCTGCGGTTTATCGGAGCTCCGATATTTTGA 596
DB 176 YKWKAYTTKRRKRWTRWTVYVYKSYMSMKTWMTAYVWTKRWKTRTKWTCTMCWK 117
QY 597 GTTGAAGAGACACTCCTTCCTTATCATCGTCGCGAAGCTGGTGAGAGATCAGTAA 656
DB 116 CTTYMAGTMYRYRYWYVYAKRAKSKRCTWTSTTTCYMKYNAKCKWSYWSMSMKWKG 57
```



```
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1664854)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g
; US-08-916-421B-1
```

Query Match 4.5%; Score 33.8; DB 4; Length 1664976;
Best Local Similarity 49.2%; Pred. No. 58;
Matches 89; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

```
QY 498 TTGTGATGGAAGAAGTCACTGAGGATATGCTGAATTTGGCTGTGGAATATGTCATAGC 557
Db 734773 TTGGTTTATTTGTGTGATGAAGTATAGTATGCTGTATGTTGTAGGTATAGGAACGGTAA 734832
QY 558 TGAGCTGCCGTTTATCGGAGCTCCGGATATTTTAGAGGTGGAAGAGACACTCCCTTGC 617
Db 734833 TGAGAGGCATTTTACAAAAGAGCTGAGGAGATTTTAAATAAAGTGGATTTATAGTGTG 734892
QY 618 TTATCATCGTCCGTGGAAGCTGGGTGAGAAGATCAGTAACCATGAATTTTCTATTTGTAT 677
Db 734893 TTATAAAATTCACAAAAGTTTGTTCAGAGGCTTAAACAGCCCAATATATACAACTGGAAT 734952
QY 678 G 678
Db 734953 G 734953
```

RESULT 12

```
US-09-692-570-1
; Sequence 1, Application US/09692570
; Patent No. 6797466
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
; Patent No. 6797466
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275C1
; CURRENT APPLICATION NUMBER: US/09/692,570
; CURRENT FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; PRIOR APPLICATION NUMBER: US 08/916,421
; PRIOR FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
```

```
;/ NAME/KEY: misc_feature
;/ LOCATION: (98266)..(98266)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (98343)..(98343)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (103998)..(103998)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (148948)..(148948)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (163385)..(163385)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (191989)..(191989)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (191995)..(191995)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (231980)..(231980)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (234187)..(234187)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (234220)..(234220)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (234814)..(234814)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (309398)..(309398)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (309418)..(309418)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (312837)..(312837)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (312993)..(312993)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (319226)..(319226)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (559167)..(559167)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (559241)..(559241)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
```

```
;/ LOCATION: (600992)..(600992)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (622708)..(622708)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (657081)..(657081)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (657203)..(657203)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (674435)..(674435)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (682442)..(682442)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (713652)..(713652)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (741684)..(741684)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (779455)..(779455)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (779676)..(779676)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (855539)..(855539)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (871619)..(871619)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (1084830)..(1084830)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (1096846)..(1096846)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (1119881)..(1119881)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (1130881)..(1130881)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (1310988)..(1310988)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (1313224)..(1313224)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (1349473)..(1349473)
```

OTHER INFORMATION: n equals a, t, c, or g

Query Match 4.5%; Score 33.8; DB 4; Length 1664976;
Best Local Similarity 49.2%; Pred. No. 58;
Matches 89; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 498 TTTCATGATCGAAGAGTCACTGAGGATATGCTGAATTTGGCTGTGGAATATGTCATAGC 557
Db 734773 TTGGTTTATTTGTGTGAAGTATAGTATGCTGTATGTTGTAGGTATAGGAAGCGGTAA 734832

QY 558 TGAGCTGCCGCTTTTATCCGAGCTCCGGATATTTAGAGGTGGAAGAGACACTCCCTTGC 617
Db 734833 TGAGAGCAATTTACAAAAGAGCTGAGAGATTTTAAATAAGTGAATTAATAGTGTG 734892

QY 618 TTATCATGCTCGGTGGAAGCTGGGTGAGAAGATCAGTAAACCATGAATTTTCTATTGTAT 677
Db 734893 TTATAAAATTTACAAAAGTTTGTGAGAGGCTTAAACAGCCAATATATACAACCTGGAAT 734952

QY 678 G 678
Db 734953 G 734953

RESULT 13

US-09-489-039A-3818
; Sequence 3818, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 3818
; LENGTH: 879
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-3818

Query Match 4.4%; Score 33.4; DB 4; Length 879;
Best Local Similarity 49.2%; Pred. No. 1.4;
Matches 88; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 473 TAATCGGCGTGCAGCGGCGTCACTTTGATGATGGAAGATCAGTGAAGATATGCTGA 532
Db 695 TCAGCGCGCGACGCGCGGAATGGGGGTGATGATCAACGACGCCCGCAGTATCTGGA 754

QY 533 ATTGGCTGTGGAATATGTCATAGCTGAGCTGCCGTTTTTTATCGAGCTCCGATATTT 592
Db 755 CCAGCGCGTGCATATGTTGCGCGGGCTGGCGCTGTTCAATCAGCGTATGCTTTTA 814

QY 593 TAGAGGTGGAAGAGACACTCTTCTTATCATCTCGTGGGAAGCTGGGTGAGAAGATC 651
Db 815 ACCTGCTGGCGAGCGCGCTCGCGATCGTCTCGATCCCCACCTGATCGCGAGCATAGC 873

RESULT 14

US-09-270-767-2403/c
; Sequence 2403, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2403

LENGTH: 337
TYPE: DNA
ORGANISM: *Drosophila melanogaster*
US-09-270-767-2403

Query Match 4.4%; Score 33.2; DB 4; Length 337;
Best Local Similarity 54.0%; Pred. No. 0.98; 58; Indels 0; Gaps 0;
Matches 68; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 93 GCAGCGCATGTTTGGTGGGATCAGCCCATTTAAACAGCAGGTTTTTCAGAGGATTATAT 152
Db 260 GCCTCGGGTGTGGAAGGTATCAACGCATTTTATATATACATCATCTTTGTT 201

QY 153 TTACAGATTAATTGGATGGCGGAAAGCTCAATTTAAAGCGTTTCTAGTTTACTTGCAGG 212
Db 200 CAAGAGATAGAGATCGAGAGGCGGCGAGTTTGGAAAAGATGTTACTTTGAAGGTGCATA 141

QY 213 GCATGA 218
Db 140 GGATAA 135

RESULT 15

US-09-270-767-17685/c
; Sequence 17685, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17685
; LENGTH: 337
; TYPE: DNA
; ORGANISM: *Drosophila melanogaster*
US-09-270-767-17685

Query Match 4.4%; Score 33.2; DB 4; Length 337;
Best Local Similarity 54.0%; Pred. No. 0.98; 58; Indels 0; Gaps 0;
Matches 68; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 93 GCAGCGCATGTTTGGTGGGATCAGCCCATTTAAACAGCAGGTTTTTCAGAGGATTATAT 152
Db 260 GCCTCGGGTGTGGAAGGTATCAACGCATTTTATATATACATCATCTTTGTT 201

QY 153 TTACAGATTAATTGGATGGCGGAAAGCTCAATTTAAAGCGTTTCTAGTTTACTTGCAGG 212
Db 200 CAAGAGATAGAGATCGAGAGGCGGCGAGTTTGGAAAAGATGTTACTTTGAAGGTGCATA 141

QY 213 GCATGA 218
Db 140 GGATAA 135

Search completed: December 10, 2004, 18:17:34
Job time : 98.1418 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2004, 16:12:37 ; Search time 447.285 Seconds
(without alignments)
9281.958 Million cell updates/sec

Title: US-10-627-124-7
Perfect score: 753
Sequence: 1 gtaagatgagatgccgaat.....aacggatcacatctgaagga 753

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4085731 seqs, 2756760397 residues

Total number of hits satisfying chosen parameters: 8171462

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 19: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
- 20: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 21: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	753	100.0	753	16	US-10-627-124-7
2	38	5.0	1194	16	US-10-282-122A-22329
3	38	5.0	1830121	14	US-10-329-960-1
4	38	5.0	1830121	16	US-10-329-960-1
5	38	5.0	1830121	18	US-10-158-865-1
6	37.4	5.0	2746	18	US-10-425-115-59495
7	36.4	4.8	945	16	US-10-425-114-29841
8	36.4	4.8	3179	16	US-10-425-114-29880
9	36.4	4.8	3285	16	US-10-425-114-11232
10	36.4	4.8	3542	16	US-10-424-599-96841
11	35.4	4.7	736	16	US-10-296-115-256
12	35.4	4.7	3681	16	US-10-424-599-6970

13	35.4	4.7	4061	10	US-09-814-353-21372	Sequence 21372, A
14	35.4	4.7	13535	10	US-09-764-891-7281	Sequence 7281, A
15	35.4	4.7	1163020	16	US-10-398-221-10	Sequence 10, Appl
16	35.4	4.7	3011208	16	US-10-398-221-2058	Sequence 2058, Ap
17	35.2	4.7	1056	18	US-10-653-047-5346	Sequence 5346, Ap
18	35.2	4.7	3117	17	US-10-437-963-10898	Sequence 10898, A
19	35	4.6	65787	16	US-10-052-482-109	Sequence 109, App
20	35	4.6	105499	17	US-10-322-281-326	Sequence 326, App
21	34.8	4.6	1355	16	US-10-425-114-32415	Sequence 32415, A
22	34.8	4.6	2509	16	US-10-425-114-24957	Sequence 24957, A
23	34.8	4.6	2791	18	US-10-425-115-63082	Sequence 63082, A
24	34.8	4.6	640681	9	US-09-790-988-1	Sequence 1, Appli
25	34.6	4.6	513	9	US-09-867-701-2391	Sequence 2391, Ap
26	34.6	4.6	526	16	US-10-424-599-92180	Sequence 92180, A
27	34.6	4.6	1984	18	US-10-881-088-33	Sequence 33, Appl
28	34.6	4.6	3931	15	US-10-006-780-1	Sequence 1, Appli
29	34.2	4.5	250	11	US-09-987-899-5354	Sequence 5354, Ap
30	34.2	4.5	1446	16	US-10-282-122A-18304	Sequence 18304, A
31	34	4.5	395	9	US-09-894-844-45	Sequence 45, Appl
32	34	4.5	385	16	US-10-388-902-45	Sequence 45, Appl
33	34	4.5	395	16	US-10-647-089-45	Sequence 45, Appl
34	34	4.5	2325	16	US-10-424-599-91791	Sequence 91791, A
35	33.8	4.5	779	13	US-10-027-632-3731	Sequence 3731, Ap
36	33.8	4.5	779	15	US-10-027-632-3731	Sequence 3731, Ap
37	33.8	4.5	875	10	US-09-814-353-20346	Sequence 20346, A
38	33.8	4.5	963	14	US-10-022-832-27	Sequence 27, Appl
39	33.8	4.5	963	16	US-10-282-122A-18554	Sequence 18554, A
40	33.8	4.5	1442	13	US-10-098-841-145	Sequence 145, Appl
41	33.8	4.5	1698	17	US-10-250-615-34	Sequence 34, Appl
42	33.8	4.5	4128	18	US-10-425-115-94844	Sequence 94844, A
43	33.8	4.5	1230025	16	US-10-289-782-1	Sequence 1, Appli
44	33.6	4.5	666	13	US-10-027-632-261431	Sequence 261431, A
45	33.6	4.5	666	15	US-10-027-632-261431	Sequence 261431, A

ALIGNMENTS

RESULT 1

US-10-627-124-7
; Sequence 7, Application US/10627124
; Publication No. US20040096944A1
; GENERAL INFORMATION:
; APPLICANT: NOVOZYMES BIOTECH, INC.
; TITLE OF INVENTION: Methods For Producing Biological Substances In Pigment-Deficient Mutants Of Bacillus Cells
; FILE REFERENCE: 10302.200-WO
; CURRENT APPLICATION NUMBER: US/10/627,124
; CURRENT FILING DATE: 2003-07-25
; PRIOR APPLICATION NUMBER: US 60/398,853
; PRIOR FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7
; LENGTH: 753
; TYPE: DNA
; ORGANISM: Bacillus subtilis
US-10-627-124-7

Query Match	100.0%;	Score	753;	DB	16;	Length	753;
Best Local Similarity	100.0%;	Pred. No.	1.4e-226;				
Matches	753;	Conservative	0;	Mismatches	0;	Indels	0;
Qy	1	GTGAATCAGATGACCGGAATGGTAAACGGAAGAGGCTGTGCATTTTATTTCGTGAGCA	60				
Db	1	GTGAATCAGATGACCGGAATGGTAAACGGAAGAGGCTGTGCATTTTATTTCGTGAGCA	60				
Qy	61	TTACAGAAAACCTGCAGAGAAATATTTGAACGGCGCAGGCATGTTTGGTGGGATCAGC	120				
Db	61	TTACAGAAAACCTGCAGAGAAATATTTGAACGGCGCAGGCATGTTTGGTGGGATCAGC	120				
Qy	121	CCATTTTACAGCAGGTTTTCAGAGGATATATTTACAGATTAATTCAGATGGCGGAAAGCT	180				

appi.001


```
NAME/KEY: misc_feature
LOCATION: (29298)..(29298)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36543)..(36543)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36551)..(36551)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36636)..(36636)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145171)..(145171)
```


; NAME/KEY: misc_feature
; LOCATION: (80024)..(80024)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (100091)..(100091)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (102696)..(102696)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (105121)..(105121)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (107248)..(107248)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (117136)..(117136)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (119750)..(119750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (119924)..(119924)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (120038)..(120038)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (121344)..(121344)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (122167)..(122167)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (122336)..(122336)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (131340)..(131340)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (131360)..(131360)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (139910)..(139910)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (140398)..(140398)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (142750)..(142750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145058)..(145058)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature

; LOCATION: (145171)..(145171)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145942)..(145942)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (147197)..(147197)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (150841)..(150841)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (152500)..(152500)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (152530)..(152530)
; FEATURE:
Query Match 5.0%; Score 38; DB 16; Length 1830121;
Best Local Similarity 55.2%; Pred. No. 57;
Matches 74; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
Qy 244 GGAACCTCCGAGCGAAAGGCTGAACGAAAGTAAGGAAAGAGGTATCACGAAACAGGAGA 303
Db 1519210 GAATTTCAAAAATGATTGGCGAAGAAACCAACGTCATAATTCGAAACGGGAAGGTCGT 1519269
Qy 304 TTTCGAGAAAGAGCCCTTGTGGCTCATGGCGGGATCCGAAAGGATTCATACATTTTCT 363
Db 1519270 TTACCTGATACAGTCATTGCTGCTGTGGCGGTGCTCGAATGCAATTGGTATGTTTGT 1519329
Qy 364 GATTTTATAGATAA 377
Db 1519330 GATTTTATGATGA 1519343
RESULT 5
US-10-158-865-1
; Sequence 1, Application US/10158865
; Publication No. US20040203093A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome, Frag
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB186P2C1D1
; CURRENT APPLICATION NUMBER: US/10/158,865
; CURRENT FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US 09/557,884
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 08/476,102
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (10150)..(10150)

```

, OTHER INFORMATION: n equals a,t,c, or g
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (29298)..(29298)
, OTHER INFORMATION: n equals a,t,c, or g
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (36543)..(36543)
, OTHER INFORMATION: n equals a,t,c, or g
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (36551)..(36551)
, OTHER INFORMATION: n equals a,t,c, or g
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (36636)..(36636)
, OTHER INFORMATION: n equals a,t,c, or g
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (40808)..(40810)
, OTHER INFORMATION: n equals a,t,c, or g
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (44416)..(44416)
, OTHER INFORMATION: n equals a,t,c, or g
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (44905)..(44905)
, OTHER INFORMATION: n equals a,t,c, or g
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (44975)..(44975)
, OTHER INFORMATION: n equals a,t,c, or g
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (45593)..(45593)
, OTHER INFORMATION: n equals a,t,c, or g
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (45732)..(45732)
, OTHER INFORMATION: n equals a,t,c, or g
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (47036)..(47036)
, OTHER INFORMATION: n equals a,t,c, or g
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (51334)..(51334)
, OTHER INFORMATION: n equals a,t,c, or g
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (51602)..(51602)
, OTHER INFORMATION: n equals a,t,c, or g
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (51786)..(51786)
, OTHER INFORMATION: n equals a,t,c, or g
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (51805)..(51805)
, OTHER INFORMATION: n equals a,t,c, or g
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (55369)..(55369)
, OTHER INFORMATION: n equals a,t,c, or g
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (65309)..(65309)
, OTHER INFORMATION: n equals a,t,c, or g
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (65313)..(65313)
, OTHER INFORMATION: n equals a,t,c, or g

```

FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (80024)..(80024)	
OTHER INFORMATION: n equals a,t,c, or g	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (100091)..(100091)	
OTHER INFORMATION: n equals a,t,c, or g	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (102696)..(102696)	
OTHER INFORMATION: n equals a,t,c, or g	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (105121)..(105121)	
OTHER INFORMATION: n equals a,t,c, or g	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (107248)..(107248)	
OTHER INFORMATION: n equals a,t,c, or g	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (117136)..(117136)	
OTHER INFORMATION: n equals a,t,c, or g	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (119750)..(119750)	
OTHER INFORMATION: n equals a,t,c, or g	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (119924)..(119924)	
OTHER INFORMATION: n equals a,t,c, or g	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (120038)..(120038)	
OTHER INFORMATION: n equals a,t,c, or g	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (121344)..(121344)	
OTHER INFORMATION: n equals a,t,c, or g	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (122167)..(122167)	
OTHER INFORMATION: n equals a,t,c, or g	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (122336)..(122336)	
OTHER INFORMATION: n equals a,t,c, or g	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (131340)..(131340)	
OTHER INFORMATION: n equals a,t,c, or g	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (131360)..(131360)	
OTHER INFORMATION: n equals a,t,c, or g	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (139910)..(139910)	
OTHER INFORMATION: n equals a,t,c, or g	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (140398)..(140398)	
OTHER INFORMATION: n equals a,t,c, or g	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (142750)..(142750)	
OTHER INFORMATION: n equals a,t,c, or g	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (145058)..(145058)	
OTHER INFORMATION: n equals a,t,c, or g	
FEATURE:	


```
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 29880
; LENGTH: 3179
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-GMROPIC021F08_FLI
US-10-425-114-29880

Query Match          4.8%; Score 36.4; DB 16; Length 3179;
Best Local Similarity 51.9%; Pred. No. 4.9;
Matches 82; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 502 ATGATGGAAGAGTCAGTCAGGAGATATGCTGAATTTGGCTGTGGAATATGTCATAGCTGAG 561
Db 2525 ATGCTGCAAGAGATGTACAATCAATGCGCCTTCTTTAGGGTCACACTTGATTTGGTGGAA 2584

QY 562 CTGCGGTTTTTTATCGGAGCTCCGGATATTTTAGAGGTGGAAGACACACTCTTGTCTTAT 621
Db 2585 ATGGTGTTCGCAAGAGATCCGAAATTTGCCGCTCTGAATGATAGACTCCTTGTTCATCA 2644

QY 622 CATCGTCCGTGGAAGCTGGGTGGAAGATCAGTAAACCA 659
Db 2645 AAGGATCTGTGGCGTTTGGGGATCAATTGAGGAACAA 2682

RESULT 9
US-10-425-114-11232
; Sequence 11232, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 11232
; LENGTH: 3285
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 701043122_FLI
US-10-425-114-11232

Query Match          4.8%; Score 36.4; DB 16; Length 3285;
Best Local Similarity 51.9%; Pred. No. 5;
Matches 82; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 502 ATGATGGAAGAGTCAGTCAGGAGATATGCTGAATTTGGCTGTGGAATATGTCATAGCTGAG 561
Db 2492 ATGCTGCAAGAGATGTACAATCAATGCGCCTTCTTTAGGGTCACACTTGATTTGGTGGAA 2551

QY 562 CTGCGGTTTTTTATCGGAGCTCCGGATATTTTAGAGGTGGAAGACACACTCTTGTCTTAT 621
Db 2552 ATGGTGTTCGCAAGAGATCCGAAATTTGCCGCTCTGAATGATAGACTCCTTGTTCATCA 2611

QY 622 CATCGTCCGTGGAAGCTGGGTGGAAGATCAGTAAACCA 659
Db 2612 AAGGATCTGTGGCGTTTGGGGATCAATTGAGGAACAA 2649

RESULT 10
US-10-424-599-96841
; Sequence 96841, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
```

```
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 96841
; LENGTH: 3542
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_58459C.1
US-10-424-599-96841
```

```
Query Match          4.8%; Score 36.4; DB 16; Length 3542;
Best Local Similarity 51.9%; Pred. No. 5.2;
Matches 82; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 502 ATGATGGAAGAGTCAGTCAGGAGATATGCTGAATTTGGCTGTGGAATATGTCATAGCTGAG 561
Db 2622 ATGCTGCAAGAGATGTACAATCAATGCGCCTTCTTTAGGGTCACACTTGATTTGGTGGAA 2681

QY 562 CTGCGGTTTTTTATCGGAGCTCCGGATATTTTAGAGGTGGAAGACACACTCTTGTCTTAT 621
Db 2682 ATGGTGTTCGCAAGAGATCCGAAATTTGCCGCTCTGAATGATAGACTCCTTGTTCATCA 2741

QY 622 CATCGTCCGTGGAAGCTGGGTGGAAGATCAGTAAACCA 659
Db 2742 AAGGATCTGTGGCGTTTGGGGATCAATTGAGGAACAA 2779
```

```
RESULT 11
US-10-296-115-256/c
; Sequence 256, Application US/10296115
; Publication No. US20040053248A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq Inc
; TITLE OF INVENTION: No. US20040053248A1e1 Nucleic Acids and Polypeptides
; FILE REFERENCE: 784PCT
; CURRENT APPLICATION NUMBER: US/10/296,115
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: US09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US09/552,317
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 1478
; SEQ ID NO 256
; LENGTH: 736
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-296-115-256
```

```
Query Match          4.7%; Score 35.4; DB 16; Length 736;
Best Local Similarity 50.9%; Pred. No. 4.4;
Matches 84; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 454 GACATGTCTCGTGAAGCGATAATCGGGCGTGC GGCGGCGTCAGTTTGTGATGATGGAAGAA 513
Db 369 GTACAGCTGTGTGAATGAGGAGCTCTTTGGGGATGGGCCCCCTTTGTAAAGTGAAGGA 310

QY 514 GTCAGTGAGGATATGCTGAATTTGGCTGTGGAATATGTCATAGCTAGCTGCCCTTTT 573
Db 309 GTCAGTCTGGGGACTGGAGGCTCTTTTGAGGATGGGGTAGCTGGGGCTCTTCTTGGGG 250

QY 574 ATCGGAGCTCCGGATATTTTAGAGGTGGAAGACACTCCTTGTCT 618
Db 249 GTTGGAAATTGCTGGGCTCTTTTAGGGGAGGGAACAGTCATAGCT 205
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2004, 13:46:32 ; Search time 159.553 Seconds
(without alignments)
7233.283 Million cell updates/sec

Title: US-10-627-124-7
Perfect score: 753
Sequence: 1 gtaagatgagatgacccgaat.....aacggatcacatctgaagga 753

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1633475 seqs, 766331085 residues

Total number of hits satisfying chosen parameters: 3266950

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents NA New:
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:
7: /cgn2_6/ptodata/2/pna/US11_NEW_COMB.seq:
8: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	346.2	46.0	4222334	1	PCT-US04-37204-1
2	345.6	45.9	750	1	PCT-US04-37204-3406
3	43.2	5.7	1101	1	PCT-US03-31318-9046
4	43.2	5.7	1457619	1	PCT-US03-31318-8739
5	38	5.0	1194	1	PCT-US02-09107B-22329
6	38	5.0	1830121	6	US-10-981-687-1
7	36.2	4.8	141753	6	US-10-746-294A-28
8	35.4	4.7	577	1	PCT-US04-35274-53829
9	35.4	4.7	600	1	PCT-US04-35274-53828
10	35.4	4.7	51879	6	US-10-990-328-94645
11	35	4.6	32943	6	US-10-990-328-94732
12	34.8	4.6	408	6	US-10-275-323A-13
13	34.4	4.6	429	6	US-10-948-737-5526
14	34.2	4.5	1446	1	PCT-US02-09107B-18304
15	34.2	4.5	499008	6	US-10-990-328-96831
16	34	4.5	198617	6	US-10-990-328-97086
17	33.8	4.5	963	1	PCT-US02-09107B-18554
18	33.8	4.5	1660	6	US-10-801-292-5
19	33.8	4.5	181323	6	US-10-990-328-93896
20	33.6	4.5	44322	6	US-10-990-328-93477
21	33.4	4.4	501	1	PCT-US02-09107B-31088
22	33.2	4.4	373600	6	US-10-990-328-93686
23	33	4.4	200	1	PCT-US03-31318-2361
24	33	4.4	28892	6	US-10-990-328-93860
25	32.8	4.4	44186	6	US-10-990-328-97821

ALIGNMENTS

RESULT 1
PCT-US04-37204-1

; Sequence 1, Application PC/TUS0437204
; GENERAL INFORMATION:
; APPLICANT: NOVOZYMES BIOTECH, INC.
; TITLE OF INVENTION: Bacillus Licheniformis Chromosome
; FILE REFERENCE: 10588.204-WO
; CURRENT APPLICATION NUMBER: PCT/US04/37204
; CURRENT FILING DATE: 2004-11-05
; PRIOR APPLICATION NUMBER: 60/535,988
; PRIOR FILING DATE: 2004-01-09
; PRIOR APPLICATION NUMBER: 60/561,059
; PRIOR FILING DATE: 2004-04-08
; PRIOR APPLICATION NUMBER: 60/572,403
; PRIOR FILING DATE: 2004-05-19
; NUMBER OF SEQ ID NOS: 8395
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4222334
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
PCT-US04-37204-1

Query Match	46.0%	Score	346.2	DB 1	Length	4222334			
Best Local Similarity	66.3%	Pred. No.	1.4e-103						
Matches	498	Conservative	0	Mismatches	253	Indels	0	Gaps	0
Qy	3	GAATGATGACCGGAAATGGTACGGAAGAGGTCTGTGTCATTTTATTTCGTGAGGCATT	62						
Db	3411675	GAATCAATGACAGAGCTTATAATGGAGACCAACACCACTATTCAAAACCGAAACTCT	3411						
Qy	63	AACAGAAACTGCAGAGAAATATTTGAACGGCGCAGGCATGTTTTGTGGGGATCAGCCC	122						
Db	3411735	TACCCAAAACCTGCAATGAAATATTTAAACGCAAGCCCATGTTCTCGTGGCATCAGCCC	3411						
Qy	123	ATTTAAACAGCAGGTTTTTCAGAGGATTATATTTACAGATTAAATTTGGATGGCGGAAAGCTCA	182						
Db	3411795	GTTTAAACGCCGATTTTCCGAAGATTATATTCATCGGCTTATCGCCTGGCGCGCTCGGTGA	3411						
Qy	183	ATTTAAACAGCGTTTTCAGTTTTTACTTCGAGGCGCATGAGCGGCTTAATCTTCTAGAAGCGCT	242						
Db	3411855	GTTTTCAGATGTATCCGTGCTTTTGGCGGAAAGGAAGCTGCCAACCTTCTCGAAGCGCT	3411						
Qy	243	TGGAATCCCGAGAGGAAGGCTGAAACGAAAGTAAAGGAAAGAGGTATCAAGAAACAGGAG	302						
Db	3411915	CGGACCCCACTATGGGAAGGCCGAAACGAAAGTTCAGAAAGAGTCTCGCGGAACCGGAG	3411						

```
QY 303 ATTTCAGAAAGACCCCTTGTGGCTCATGCGGGATCCGAAGCGATTTCATACATTTTC 362
DB 3411975 ATTTCGTAAAGCGCTTGAAGCGCATGCGGAAATCCCGAGGACATCCATACATTTTC 3412034
QY 363 TGATTTTATAGATAACAAAGCCCTACCAAGCTGTTTGAGACAAGAAAGTTGAACATCATTTT 422
DB 3412035 CGATTTTCGGCAACAGAGACCGCATACCGGAATTTTGGGATGGAAGTGAAGCTCGCTTTT 3412094
QY 423 TGAGCAGCTCATTTTCGACATGCTTTGTTGGACATGCTCTGGAAGCGGATATTCGGGG 482
DB 3412095 CGACCAAGACGCAATTTTGGCAATGCTGCTCGGAGATGTCCGATGCGGCTATCTCCGAGC 3412154
QY 483 TGCGCGGGGCGTCAGTTTGTATGATGGAAGAAAGTCAGTGAGATATGCTGAATTTGGCTGT 542
DB 3412155 GGCGCGGGGCACTCGGATGATGCTGGAAGTTCAGCGGAGACATGCTGGAGCTGGCTGT 3412214
QY 543 GGAATATGTCATAGCTGAGCTGCCGTTTTTTTATCGGAGCTCCGGAATATTTTAGAGTGA 602
DB 3412215 TGAATAGTTCATCGTGAACCTTCGTTTTTTCATCGCGGCCCTGATATTTTAGCGGTGA 3412274
QY 603 AGAGACACTCTGCTTATCATGCTCGTCCGTAAGCTGGGTGAGAAGATCAGTAACCATGA 662
DB 3412275 AGAGACGCTTCTGCTTATCACCAGGCAATGGAAGCTCGGCAACAGATCTCCCGTAATGA 3412334
QY 663 ATTTCCTATTGTATGCGCGCAATCAAGGCTATCTCATTTGTACAGGAATGCGGAGAT 722
DB 3412335 ATTTCGCGTCAAAATGGCGCGCAATCAAGGATATCTCATGTTTTCGAAGCGGACGAAG 3412394
QY 723 GCTTTCTGAGAAACGGATCATCTCTGAAGGA 753
DB 3412395 GGTGGAATCTAAGCATGAGGGAACGA 3412425
```

RESULT 2

```
PCT-US04-37204-3406
; Sequence 3406, Application PC/TUS0437204
; GENERAL INFORMATION:
; APPLICANT: NOVOZYMES BIOTECH, INC.
; APPLICANT: NOVOZYMES A/S
; TITLE OF INVENTION: Bacillus Licheniformis Chromosome
; FILE REFERENCE: 10588.204-WO
; CURRENT APPLICATION NUMBER: PCT/US04/37204
; CURRENT FILING DATE: 2004-11-05
; PRIOR FILING DATE: 2004-01-09
; PRIOR FILING DATE: 2004-04-08
; PRIOR FILING DATE: 2004-04-08
; PRIOR FILING DATE: 2004-05-19
; NUMBER OF SEQ ID NOS: 8395
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3406
; LENGTH: 750
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
PCT-US04-37204-3406
```

```
Query Match 45.9%; Score 345.6; DB 1; Length 750;
Best Local Similarity 66.5%; Pred. No. 1.7e-105;
Matches 495; Conservative 0; Mismatches 249; Indels 0; Gaps 0;
```

```
QY 10 ATGACCGAATGTTAAACGGAAGAGTCTGTGCATTTTATTGCTGAGGCATTAAACAGAA 69
DB 1 ATGACAGAGCTTAAATGGGAGAGCAACACAGGCTATTCAAAACCGAAATCTTTACCCAA 60
QY 70 AACTGCAGAGAAATATTGTAACGCGCAGGAGCATGTTTGTGGGATCAGCCCATTTAAC 129
DB 61 AACTGCATGAAATATTAAACGAGAGCGCAATGTTCTGTCGCAATCAGCCGTTTAAAC 120
QY 130 AGCAGGTTTTTCAGAGGATTTATTTACAGATTAATTCGATGGGCAAGAGCTCAATTTAA 189
DB 121 AGCCGATTTTCCGAAGATTTATATTTCATCGGCTTATCGCTGGGCGTCCGCTGAGTTTCAG 180
```

```
QY 190 AGCGTTTCAGTTTACTTTCAGGGCATGAGGGCTTAATCTTCTAGAAGCGCTTGAAC 249
DB 181 AGTGATTCGTGCTTTTGGCGGGAAGAGCTGCCAACCTTCTCGAAGCGCTCGCAC 240
QY 250 CCGAGAGAAAGCGCTGAACGAAAGATGAAGAAAGGTATCAGAAACAGGAGATTGCA 309
DB 241 CCACATGGGAAGCGCGCAACGAAAGTCAAGAAAGTCTCGCGAAACCGGAGATTGCT 300
QY 310 GAAAGAGCCTTTGTGCTCATGCGGGGATCCGAAGCGGATTCATACATTTTCTGATTTT 369
DB 301 GAAAGAGCCTTTGGAAGCGCATGCGGAAATCCCGAGGACATCCATACATTTTCCGATTT 360
QY 370 ATAGATAACAAAGACCTTACCAGCTGTTTGAGACAAGAAAGTTGAAACATGCATTTTGAAGCAG 429
DB 361 GCGAACACGACCGCATACCGGAATTTTGGGATGGAAGTGAAGCTGCTTTTTCGACCAG 420
QY 430 CTTCAATTTTCGACATGCTTTGTTGACATGCTCTGTTGAAGCGGATAATTCGGGCGTGC 489
DB 421 AGCAATTTTCGCAATGCTTGCCTGGAGATGTGCGATGCGGCTATCTCGAGCGGCGCG 480
QY 490 GCGCTCAGTTTGTATGATGGAAGAGTCACTGAGGATATGCTGAATTTGGCTGTGGAATAT 549
DB 481 GGCATCTCGATGGAATGCTGTGAAGTCAGCGGAGACATGCTGGAGCTGGCTGTTGAATAC 540
QY 550 GTCATAGCTGAGCTGCCGTTTTTTTATCGGAGCTCCGATATTTTAGAGGTGGAAGACGA 609
DB 541 GTCATGCTGAACTTCGTTTTTTCATCGCGGCCCTGATATTTTAGGCGTCAAGAGACG 600
QY 610 CTCCTTGTCTTATCATCGTCCGTGGAAGCTGGGTGGAAGATCAGTAACCATGAAATTTTCT 669
DB 601 CTTCTTGTCTTATCAGCGGCATGGAAGCTCGGCAACAGATCTCCGTAATGAATTTGCC 660
QY 670 ATTTGTATGCGCGCAATCAAGGATATCTCATTTGTACAGAAATGCGGAGATGCTTCT 729
DB 661 GTCAAAATCGCGCAATCAAGGATATCTCATGTTTTCGAAGCGGACGAAGGTTGGA 720
QY 730 GAGAAACGGATCACATCTGAAGGA 753
DB 721 TCTAAAAGCATGCAAGGAGNACGA 744
```

RESULT 3

```
PCT-US03-31318-9046/c
; Sequence 9046, Application PC/TUS0331318
; GENERAL INFORMATION:
; APPLICANT: Regence of the University of Minnesota
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
; FILE REFERENCE: 09531-128WO1
; CURRENT APPLICATION NUMBER: PCT/US03/31318
; CURRENT FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: US 60/416,395
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 11433
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9046
; LENGTH: 1101
; TYPE: DNA
; ORGANISM: Lawsonia intracellularis
PCT-US03-31318-9046
```

```
Query Match 5.7%; Score 43.2; DB 1; Length 1101;
Best Local Similarity 60.0%; Pred. No. 0.00088;
Matches 72; Conservative 0; Mismatches 48; Indels 0; Gaps 0;
```

```
QY 350 TTCAATCATTTCTTGATTTTATAGATAACAAAGCCTACCAAGCTGTTGAGACAAGAGTTG 409
DB 561 TTCAATTTTCTCTCAATTAATGATGAGGTTTCCACTATTTAAAGTGAAGAAAG 502
QY 410 AACATGATTTTTTGAAGCGCTCATTTTCGACATGCTTGTGGAATGCTCTCGTGAAG 469
DB 501 AAGATTCATTTCTCAACGCCCACTTTAGCGTCTTTATTATTTTTCCTATTTTGAAG 442
```

```

RESULT 4
PCT-US03-31318-8739
; Sequence 8739, Application PC/TUS0331318
; GENERAL INFORMATION:
; APPLICANT: Regents of the University of Minnesota
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
; FILE REFERENCE: 09531-128W01
; CURRENT APPLICATION NUMBER: PCT/US03/31318
; PRIOR FILING DATE: 2003-10-01
; PRIOR FILING DATE: 2003-10-01
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 11433
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 8739
; LENGTH: 1457619
; TYPE: DNA
; ORGANISM: Lawsonia intracellularis
PCT-US03-31318-8739

Query Match          5.7%; Score 43.2; DB 1; Length 1457619;
Best Local Similarity 60.0%; Pred. No. 0.051;
Matches 72; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 350 TTGATGATTTTCTGATTTTATAGATAACAAAGCCTACAGCTGTTGAGACAAAGATTG 409
Db 184520 TTGATTTTCTCATTAATAGTAGATGCAGGGTTTCCACTATTAAAGTGAAGAAAG 184579

Qy 410 AACATGATTTTTCAGACGCTCATTTTCGACATGCTTCTTGGACATCTCTCGTGAAG 469
Db 184580 AAGATTCATTCTTCAACAGCCACTTTAGCGTCTTTATTATTTTTCGCTATTTTGAAG 184639

RESULT 5
PCT-US02-09107B-22329
; Sequence 22329, Application PC/TUS0209107B
; GENERAL INFORMATION:
; APPLICANT: Elittra Pharmaceuticals Inc.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034VPC
; CURRENT APPLICATION NUMBER: PCT/US02/09107B
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 09/948,993
; PRIOR FILING DATE: 2001-09-06
; PRIOR APPLICATION NUMBER: 60/342,923
; PRIOR FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 10/072,851
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: 60/362,699
; PRIOR FILING DATE: 2002-03-06
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22329
; LENGTH: 11194
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
PCT-US02-09107B-22329

Query Match          5.0%; Score 38; DB 1; Length 11194;
Best Local Similarity 55.2%; Pred. No. 0.052;
Matches 74; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Qy 244 GGAATCTCGAGAGAAAGCGTGAACGAAAGTAGAGAAAGAGGTATCAGAAACAGAGA 303
Db 610 GAATTTCAAAAATGATGCGGAGAAACCAAAAGTCAAAATCTCGAACGGAAGTCTG 669

Qy 304 TTTGCGAAGAGCCCTTGTGGTTCATGCGGGGATCCGAGGCGATTCATACATTTTCT 363
Db 670 TTACCTGATACGATCATGCTGCTGTTGGCGGTGGCTCGAATGCAATGTTATGTTGCT 729
```

```

Qy 364 GATTTATAGATAA 377
Db 730 GATTTATTGATGA 743

RESULT 6
US-10-981-687-1
; Sequence 1, Application US/10981687
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome, Frag
; FILE REFERENCE: PB186P2C1D12
; CURRENT APPLICATION NUMBER: US/10/981,687
; CURRENT FILING DATE: 2004-11-05
; PRIOR APPLICATION NUMBER: US 10/158,865
; PRIOR FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US 09/557,884
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 08/476,102
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (40808)..(40810)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44416)..(44416)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44905)..(44905)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a,t,c, or g
```

```
FEATURE:
NAME/KEY: misc feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
```

```
NAME/KEY: misc feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
```

Query Match 5.0%; Score 38; DB 6; Length 1830121;
Best Local Similarity 55.2%; Pred. No. 3.2;
Matches 74; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Qy	244	GGAACTCCGAGAGGCTGAACGAAAGTAAGGAAAGAGGTATCCGAAACAGGAGA	303
Db	1519210	GAATTTCAAAATGATTGCGAAGAACCAACAGTCAATTCGAAACGGAAGTCT	1519269
Qy	304	TTTGCAGAAAGACCCCTTGTGGCTCATGCGGGGATCCGAGGCGATTACATTTCT	363
Db	1519270	TTACCTGATACAGTCATTGCTGTGGCGGTGCTCGAATGCAATTGGTATGTTGCT	1519329


```
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING HUMAN ENZYME PROTEINS, METHODS OF DETECTION AND
; FILE REFERENCE: CL001495
; CURRENT APPLICATION NUMBER: US/10/990,328
; CURRENT FILING DATE: 2004-11-17
; NUMBER OF SEQ ID NOS: 58824
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 94645
; LENGTH: 51879
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-990-328-94645

Query Match          4.7%; Score 35.4; DB 6; Length 51879;
Best Local Similarity 50.9%; Pred. No. 3.2;
Matches 84; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 454 GACATGCTCTGTAAGCGAATAATCGGGCGTGGGGGGCTCAGTTTGATGATGGAAGAA 513
Db 41590 GTCACAGCTGTGGAATGAGGAGCTCTTTGGGGGATGGGGCCCTTTGTAAAGTGAAGGA 41531

QY 514 GTCAGTGAGGATATGCTGAATTTGGCTGTGGAATATGTCATAGCTGAGCTGCCGTTT 573
Db 41530 GTCACGCTGGGAGCTGGAGGCTCTTTGAGGATGGGTAGCTGGGGCTTCTTTGGGG 41471

QY 574 ATCGGAGCTCCGGATATTTTAGAGGTGGAAGAGACACTCTTGTCT 618
Db 41470 GTTGAATTCTGGGGTCTTTTATGGGAGGGAACAGTCATAGCT 41426

RESULT 11
US-10-990-328-94732
; Sequence 94732, Application US/10990328
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING HUMAN ENZYME PROTEINS, METHODS OF DETECTION AND
; FILE REFERENCE: CL001495
; CURRENT APPLICATION NUMBER: US/10/990,328
; CURRENT FILING DATE: 2004-11-17
; NUMBER OF SEQ ID NOS: 58824
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 94732
; LENGTH: 32943
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(32943)
; OTHER INFORMATION: n = A,T,C or G
US-10-990-328-94732

Query Match          4.6%; Score 35; DB 6; Length 32943;
Best Local Similarity 49.7%; Pred. No. 3.4;
Matches 89; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 30 AAGAAGTCTGTGCATTTTATGCTGAGGCAATTAACAGAAAACGACAGAGAAATATTGA 89
Db 21106 AGAAAGTACTTACATTCCTCATTTAACAGAGAGAAATATGTAAGAGTATTGT 21165

QY 90 ACGCGCAGGCGATGTTTTGGTGGGATCAGCCCATTTTAAACGACAGTGTTCAGAGGATTA 149
Db 21166 CCATGTCACACAGTTAGTGGTGGCAGAACTGCGATTCAAACCCAGTGTTCAGAGACTGC 21225

QY 150 TATTACAGATTAATGATGGGCGGAAGCTCAATTTAAAGCGTTTCAGTTTACTTG 208
Db 21226 TTTTAAACCACTACATCTTCTGTCCCTCTAATCATGAAATTAATGCAACTTTTACTTIG 21284
```

```
RESULT 12
US-10-275-323A-13/c
; Sequence 13, Application US/10275323A
; GENERAL INFORMATION:
; APPLICANT: DONNE-GOUSSE, CAROLE
; APPLICANT: LAUDET, VINCENT
; APPLICANT: HANNI, CATHERINE
; TITLE OF INVENTION: METHOD OF DETECTING AND IDENTIFYING THE PRESENCE OF BIOLOGICAL
; TITLE OF INVENTION: MATERIALS DERIVED FROM BIRDS, AND OLIGONUCLEOTIDES FOR CARRYING
; TITLE OF INVENTION: IT OUT
; FILE REFERENCE: 0508-1052
; CURRENT APPLICATION NUMBER: US/10/275,323A
; CURRENT FILING DATE: 2002-11-08
; PRIOR APPLICATION NUMBER: PCT/FR01/01279
; PRIOR FILING DATE: 2001-04-26
; PRIOR APPLICATION NUMBER: FR 00/05850
; PRIOR FILING DATE: 2000-05-09
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 408
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DNA fragment amplified
; OTHER INFORMATION: from the genome of birds
; NAME/KEY: modified_base
; LOCATION: (43)..(43)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (57)..(57)
; OTHER INFORMATION: a, c, g, or t
; NAME/KEY: modified_base
; LOCATION: (71)..(71)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (134)..(134)
; OTHER INFORMATION: a, c, g, or t
; NAME/KEY: modified_base
; LOCATION: (212)..(212)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (214)..(214)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (217)..(217)
; OTHER INFORMATION: a, c, g, or t
; NAME/KEY: modified_base
; LOCATION: (226)..(227)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (236)..(236)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (245)..(245)
; OTHER INFORMATION: a, c, g, or t
; NAME/KEY: modified_base
; LOCATION: (253)..(253)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (269)..(269)
```

OTHER INFORMATION: a, c, g, or t
FEATURE:
NAME/KEY: modified base
LOCATION: (324)..(324)
OTHER INFORMATION: a, c, g, or t
FEATURE:
NAME/KEY: modified base
LOCATION: (330)..(330)
OTHER INFORMATION: a, c, g, or t
FEATURE:
NAME/KEY: modified base
LOCATION: (352)..(352)
OTHER INFORMATION: a, c, g, or t
FEATURE:
NAME/KEY: modified base
LOCATION: (355)..(355)
OTHER INFORMATION: a, c, g, or t
FEATURE:
NAME/KEY: modified base
LOCATION: (357)..(357)
OTHER INFORMATION: a, c, g, or t
FEATURE:
NAME/KEY: modified base
LOCATION: (362)..(362)
OTHER INFORMATION: a, c, g, or t
FEATURE:
NAME/KEY: modified base
LOCATION: (364)..(364)
OTHER INFORMATION: a, c, g, or t
FEATURE:
NAME/KEY: modified base
LOCATION: (406)..(406)
OTHER INFORMATION: a, c, g, or t
US-10-275-323A-13

Query Match 4.6%; Score 34.8; DB 6; Length 408;

Best Local Similarity 8.9%; Pred. No. 0.33; Mismatches 157; Indels 0; Gaps 0;

Matches 35; Conservative 200;

QY 92 GCGCGGCGGATGTTTGGTGGGATCAGCCATTAAACAGCAGGTTTTCAGAGGATTATA 151
DB 403 KRWRSWAKGHRYYRDBDKKYVYRYYRDRDHYRDBVNDKTKNRRANDRDRDT 344
QY 152 TTTACAGATTAAATGGATGGCGAAAGCTCAATTAAAGGTTTTCAGTTTACTTCGAG 211
DB 343 WRTRKDDWYNNRDKTDNDWBHBBYTRGTRCTGCGMTYCTMTAATAADKWTREDH 284
QY 212 GCGATGAGCGGCTAATCTCTTAGAAGCGCTTGGAATCCGAGAGAAAGGCTGAACGAA 271
DB 283 BRHDHRTKHDWYNNKWHGKHVRVNDWDBVNNRDRWRSSNRNWTWDDRRNDG 224
QY 272 AAGTAAGGAAGAGGTATCAGAAACAGGAGATTTCGAGAAAGAGCCCTTGTGCTCATG 331
DB 223 VRRTWDRNDNDNDNDWTDVDDKDDDDKDDDDKDDDDKDDDDKDDDDKDDDDKDDDD 164
QY 332 GCGGCGATCCGAGCGGATTCATACATTTTCTGATTTTATAGATAACAAAGCCCTACCAGC 391
DB 163 WDAWYDGHKAARWHYDTGTTKHTTYNKDDKYSYRDRRRWRYRSRDRRDKWD 104
QY 392 TGTGAGACAAGAGTTGAACATGCTTTTGTAGAGCGCTCATTTTCGACATGCTTGT 451
DB 103 WTKKWTWRHRTATGCTCCDRAGCATTCACYNRRYARCAVYBWDVNVNTRGRDBDGYGA 44
QY 452 TGGACATGCTCGTGAAGCGATATCGGCGT 483
DB 43 NRKRCMWAACYRAATGHRHMAAGTGCA 12

RESULT 13

US-10-948-737-5526

; Sequence 5526, Application US/10948737

; GENERAL INFORMATION:

; APPLICANT: CHAN, VIVIEN W.

APPLICANT: ESCOBEDO, JAIME
APPLICANT: GARCIA, PABLO DOMINGUEZ
APPLICANT: HANSEN, RHONDA
APPLICANT: KAUFMANN, JOERG
APPLICANT: KENNEDY, GIULIA C.
APPLICANT: LAMSON, GEORGE
APPLICANT: MOLER, EDWARD J.
APPLICANT: RANDAZZO, FILIPPO
APPLICANT: REINHARD, CHRISTOPH
APPLICANT: SUDDUTH-KLINGER, JULIE
TITLE OF INVENTION: GENE PRODUCTS DIFFERENTIALLY EXPRESSED
FILE REFERENCE: 2300-21987
CURRENT APPLICATION NUMBER: US/10/948,737
CURRENT FILING DATE: 2004-09-22
PRIOR APPLICATION NUMBER: 10/616,900
PRIOR FILING DATE: 2003-07-09
PRIOR APPLICATION NUMBER: 09/872,850
PRIOR FILING DATE: 2001-06-01
PRIOR APPLICATION NUMBER: 60/208,871
PRIOR FILING DATE: 2000-06-02
PRIOR APPLICATION NUMBER: 10/081,519
PRIOR FILING DATE: 2002-02-21
PRIOR APPLICATION NUMBER: 60/270,959
PRIOR FILING DATE: 2001-02-21
PRIOR APPLICATION NUMBER: 10/310,673
PRIOR FILING DATE: 2003-01-08
PRIOR APPLICATION NUMBER: 60/336,613
PRIOR FILING DATE: 2001-12-04
PRIOR APPLICATION NUMBER: US03/00657
PRIOR FILING DATE: 2003-01-08
PRIOR APPLICATION NUMBER: 60/345,637
PRIOR FILING DATE: 2002-01-08
PRIOR APPLICATION NUMBER: 10/081,124
PRIOR FILING DATE: 2002-02-21
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 13996
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5526
LENGTH: 429
TYPE: DNA
ORGANISM: Homo sapiens
US-10-948-737-5526

Query Match 4.6%; Score 34.4; DB 6; Length 429;

Best Local Similarity 52.9%; Pred. No. 0.47;

Matches 74; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 423 TGAGGAGCCTCATTTTCGACATGCTGTTTGGACATGCTCTCGTGAACGATATCGGGCG 482
DB 253 TGAGGACCACTTTTGGAGCTCATTTAATGGTCAAGAGATGAAGTGGATGCTGCGGTG 312
QY 483 TGCGCGGGCGTCACTTGTGATGATGAGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 542
DB 313 GACTGGAGAGGCAAGTGTGTAAGGAGCGAGTGTGTTGATGATGATGATGATGATGATGAT 372
QY 543 GGAATATGTCATAGCTCAGC 562
DB 373 CAAAAGTTTAGATTAAAGC 392

RESULT 14

PCT-US02/09107B-18304

; Sequence 18304, Application PC/TUS0209107B

; GENERAL INFORMATION:

; APPLICANT: Elitra Pharmaceuticals Inc.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034VPC

; CURRENT APPLICATION NUMBER: PCT/US02/09107B

; CURRENT FILING DATE: 2002-03-12

; PRIOR APPLICATION NUMBER: 09/815,242

; PRIOR FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 09/948,993

Job time : 182.553 secs

```
; PRIOR FILING DATE: 2001-09-06
; PRIOR APPLICATION NUMBER: 60/342,923
; PRIOR FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 10/072,851
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: 60/362,699
; PRIOR FILING DATE: 2002-03-06
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 18304
; LENGTH: 1446
; TYPE: DNA
; ORGANISM: Campylobacter jejuni
PCT-US02-09107B-18304

Query Match      4.5%; Score 34.2; DB 1; Length 1446;
Best Local Similarity 49.7%; Pred.No.1.1;
Matches 87; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

Qy      206  TTCGAGGCGCATGAGCGGCTAATCTTCTAGAGCGCTTGGAACCTCCGAGAGGAAAGGCTG 265
Db      92  TTGCAGTTTTTACAGAGCCCTGATCGTGAATGTTTGCATGTAATAATCGCAGATGAGCTT 151

Qy      266  AACGAAAAAGTAAGAAAGAGGTATCACGAAACAGAGGATTTGCAGAAAGAGCCCTTGTGG 325
Db     152  ATCGCATAGGACGGATGCTATAAGAGGATATTTAGATGTTGCTCGTATTGTCGAGATTG 211

Qy      326  CTGATGGCGGGATCCGAAAGGATTCATACATTTCTGATTTTATAGATAACAA 380
Db     212  CTAAGGCTTGTGTGCTGATGCGATTCATCCTGTTATGGATTTTAAAGTGAGAA 266

RESULT 15
US-10-990-328-96831/c
; Sequence 96831, Application US/10990328
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING HUMAN ENZYME PROTEINS, METHODS OF DETECTION AND
; FILE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001495
; CURRENT APPLICATION NUMBER: US/10/990,328
; CURRENT FILING DATE: 2004-11-17
; NUMBER OF SEQ ID NOS: 558824
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 96831
; LENGTH: 499008
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(499008)
; OTHER INFORMATION: n = A,T,C or G
US-10-990-328-96831

Query Match      4.5%; Score 34.2; DB 6; Length 499008;
Best Local Similarity 52.4%; Pred.No.29;
Matches 75; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy      158  GATTAATTTGGATGGCGGAAGCTCAATTTAAAGCGTTTCAGTTTCTTACTTGCAGGGCATG 217
Db     129915  GCTTGGGTGGAGGTGGGTGAGTACATAGAAAGGGCGCAAGAAACAAAGAGAGGTAGTG 129856

Qy      218  AGCGGCTAATCTTCTAGAGCGCTTGGAACCTCCGAGAGGAAAGGCTGAAACGAAAGTAA 277
Db     129855  TGAAGATAAGAGGTTAGACAAAGCAGAGAACCCCTTAGGAGAGGGATCAGCATATAGAAA 129796

Qy      278  GGAAGAGGTATCACGAACAGG 300
Db     129795  GGAATGAAGTATGAGGAGCATG 129773

Search completed: December 10, 2004, 20:51:35
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2004, 13:18:28 ; Search time 2775.16 Seconds
(without alignments)
9887.411 Million cell updates/sec

Title: US-10-627-124-7
Perfect score: 753
Sequence: 1 gtaagagatgacccgaat.....aacggatcacatctgaaga 753

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	43.2	5.7	585	8	BZ484107
C 2	43.2	5.7	818	8	BH431961
C 3	43.2	5.7	825	8	BH55688
C 4	43.2	5.7	844	8	BZ429521
C 5	43.2	5.7	916	9	CNS005MM
C 6	40	5.3	1013	9	CL053107
C 7	40	5.3	1101	9	CNS0039G
C 8	39.6	5.3	414	8	AZ001661
C 9	39.6	5.3	498	8	AZ289765
C 10	39.6	5.3	552	8	AZ511287
C 11	39.6	5.3	717	9	AG517041
C 12	39.4	5.2	309	7	CN215521
C 13	39	5.2	926	7	CR442585
C 14	39	5.2	1101	9	CNS017KX
C 15	38.4	5.1	1021	9	CNS00DL5
C 16	38.4	5.1	1103	9	CL082028
C 17	38.2	5.1	1098	8	CC187205
C 18	38	5.0	555	2	AW705941
C 19	37.8	5.0	802	5	BK723518
C 20	37.6	5.0	800	8	BZ470290
C 21	37.6	5.0	1101	9	CNS0100X
C 22	37.4	5.0	491	1	AA516933
C 23	37.4	5.0	505	8	AQ721360
C 24	37.4	5.0	543	6	CB929916

25	37.4	5.0	589	2	AW032477	AW032477 EST276036
C 26	37.4	5.0	703	8	BH936464	BH936464 O960603.
C 27	37.2	4.9	620	5	BQ527323	BQ527323 NISC no21
C 28	37.2	4.9	1101	9	CNS00LOO	AL068607 Drosophila
C 29	37	4.9	466	9	CG470635	CG470635 ZMMBB5027
C 30	37	4.9	766	8	AZ600588	AZ600588 LM0418B05
C 31	36.8	4.9	307	9	CL571383	CL571383 OB_BA001
C 32	36.8	4.9	414	8	AQ064406	AQ064406 HS_2198B
C 33	36.8	4.9	480	9	CL602544	CL602544 OB_BA007
C 34	36.8	4.9	531	9	CL582438	CL582438 OB_BA004
C 35	36.8	4.9	570	4	B1785498	B1785498 sai1f11.
C 36	36.8	4.9	733	8	BZ028160	BZ028160 oeg78d08.
C 37	36.8	4.9	960	5	EX442989	EX442989 BX442989
C 38	36.8	4.9	1070	8	BH655582	BH655582 BOMNCL1TR
C 39	36.6	4.9	676	7	CO090490	CO090490 GR_BA101
C 40	36.4	4.8	387	5	BH050997	BH050997 1111037B0
C 41	36.4	4.8	494	2	BE803001	BE803001 sr47f05.Y
C 42	36.4	4.8	553	4	B1973129	B1973129 saf40c05.
C 43	36.4	4.8	559	1	AV826780	AV826780 AV826780
C 44	36.4	4.8	587	9	CC962904	CC962904 BOIBJ93TF
C 45	36.4	4.8	592	4	B1974352	B1974352 sai99e11.

ALIGNMENTS

RESULT 1
LOCUS BZ484107/c 585 bp DNA linear GSS 13-DEC-2002
DEFINITION BONKL22TF BO_1.6_2_KB tot Brassica oleracea genomic clone BONKL22,
genomic survey sequence.
ACCESSION BZ484107
VERSION BZ484107.1 GI:26786505
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
REFERENCE 1 (bases 1 to 585)
AUTHORS Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other GSSs: BONKL22TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.

FEATURES

source
Location/Qualifiers
1..585
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BONKL22"
/clone_lib="BO_1.6_2_KB tot"
/notes="Vector: pHO51; Site 1: BstXI; 1.6-2 kb sheared
total DNA inserted into pHO51 using BstXI linkers"

ORIGIN

Query Match 5.7%; Score 43.2; DB 8; Length 585;
Best Local Similarity 50.0%; Pred. No. 0.23;
Matches 108; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
Qy 103 GTTTTGGTGGGATCCACCATTTTACAGCAGGTTTTCAGAGGATTATTATTCAGATTATTA 162
Db 504 GTTGTGTCAGGAACATGCTTCACCCAGGAATCATCCAAAGGAGAAAGCAGAGATGC 445

AUTHORS Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of *Brassica oleracea*
JOURNAL Unpublished (2001)
COMMENT Other GSSs: BOOAY61TF
 Contact: Chris Town
 TIGR Medical Center Drive, Rockville, MD 20850, USA.
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TR
 Class: sheared ends.

Location/Qualifiers
 1. .844
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="TOL000DH3"
 /db_xref="taxon:3712"
 /clone="BOOAY61"
 /note="Vector: pHO51; Site 1: BstXI; 1.6-2 kb sheared total DNA inserted into pHO51 using BstXI linkers"

FEATURES

source

ORIGIN

Query Match 5.7%; Score 43.2; DB 8; Length 844;
 Best Local Similarity 50.0%; Pred. No. 0.25;
 Matches 108; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 103 GTTTTGGTGGGATCAGCCCAATTAACAGCAGGTTTCAGAGGATTAATATTCAGATTAA 162
 DB 39 GTTTTGGTGGGATCAGCCCAATTAACAGCAGGTTTCAGAGGATTAATATTCAGATTAA 162
 QY 163 ATTGGATGGCGAAGCTCAATTTAAAGCGTTTCAGTTTACTTTCAGAGGATTAATATTCAGATTAA 162
 DB 99 AGTGAAGAGCTTCAGCTGACTGTCAAGGGGTTGAAGATGGAGATGAAGATGAAGAG 158
 QY 223 GCTAATCTTCAGAGCGCTTGGAACTCCGAGAGGAAAGGCTGAACGAAAGTAAGGAAA 282
 DB 159 GGAACATGTGTGCAGATGAAGAAAGCAAAAGGAAGAGCTGATGAAGAACTGTGTAGTAA 218
 QY 283 GAGGTATCAGAAACAGGAGATTTGCAGAAAGGCC 318
 DB 219 AAAGAGCGAGAAACTCCAGAGAGAGGCAAGAGAC 254

RESULT 5

CNS005MM

LOCUS

DEFINITION CNS005MM 916 bp DNA linear GSS 03-JUN-1999
 Drosophila melanogaster genome survey sequence T7 end of BAC #
 BACR12016 of RPCI-98 library from *Drosophila melanogaster* (fruit fly); genomic survey sequence.

ACCESSION

AL059388

VERSION

AL059388.1

KEYWORDS

GSS.

SOURCE

ORGANISM

Drosophila melanogaster (fruit fly)

Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 916)

Genoscope.

Direct Submission

Submitted (02-JUN-1999)

Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr)

Web : www.genoscope.cns.fr

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the *Drosophila* melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP *Drosophila* melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mammose in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's Pi and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

Location/Qualifiers

1. .916
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /clone="BACR12016"
 /clone_lib="RPCI-98"
 /note="end : T7"

ORIGIN

Query Match 5.7%; Score 43.2; DB 9; Length 916;
 Best Local Similarity 39.2%; Pred. No. 0.26;
 Matches 102; Conservative 17; Mismatches 141; Indels 0; Gaps 0;

QY 418 TTTTTCAGCAGCCTCATTTTCGACATGCTTTGACATGCTCTCGTGAAGCATATC 477
 DB 232 TTTTTCAGCAGCCTCATTTTCGACATGCTTTGACATGCTCTCGTGAAGCATATC 477
 QY 478 GGGCGTGGCGGGCGTCAAGTTGATGATGGAAGAGTCAGTGAAGATGCTGAATTTG 537
 DB 292 NCGGTGKGGCTTKNSYTTTTTGGTTTWTGAGTTTTTGTGTTTGTGTTGTTG 351
 QY 538 GCTGTGAATATGATGATGCTGAGTGCCTTTTTTATCGAGCTCCGGATATTTAGAG 597
 DB 352 GNKTTGGTGTGCTTTTTTTCACGNCAAATTTTTCACGCGTNTGCAANTTTT 411
 QY 598 GTGGAAGAGACATCCCTGCTTATCATCGTCCGCGGAGCTGGTGAAGATGAGTAAC 657
 DB 412 CCNTCAGAAGCATTTTTCNTTCATTTTACNYCGNNYGAACNYGAGATTCNTAAG 471
 QY 658 CATGAATTTTCTATTTGTAT 677
 DB 472 WCAATTTTANYNTTGTANT 491

RESULT 6

LOCUS

CL053107/c

DEFINITION

CL053107

ACCESSION

CL053107

VERSION

CL053107.1

KEYWORDS

GSS.

SOURCE

ORGANISM

Xenopus tropicalis (western clawed frog)

Xenopus tropicalis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopodinae; Xenopus; Silurana.

1 (bases 1 to 1013)

Kremizki, C., Carter, J., McPherson, J., Warren, W., Graves, T., Mardis, E. and Wilson, R.

A physical map of the xenopus tropicalis genome

Unpublished (2003)

Contact: Richard K Wilson

Genome Sequencing Center

Washington University School of Medicine

Email: submissions@watson.wustl.edu

Insert Length: 175000 Std Error: 0.00

Seq primer: RM1 TAGCACTCATATAGGAGA

Class: BAC ends

High quality sequence start: 87

High quality sequence stop: 526.

Location/Qualifiers

1. .1013

/organism="Xenopus tropicalis"


```

QY 373 GATAACAAAGCTACCGAGCTGTTTCAGACAGAGATTGCAACATGCA 417
Db 118 GCAACAGATCCGATTCAGAGTCAGATCCAGAAAGTTGAGATGCA 74

RESULT 13
CR442585/c
LOCUS DEFINITION
CR442585 926 bp mRNA linear EST 19-JUN-2004
CR442585 XGC-tailbud Xenopus tropicalis cDNA clone TTbA073h24 3',
mRNA sequence.
CR442585.1 GI:48968172
CR442585 EST.
SOURCE
ORGANISM
Xenopus tropicalis (western clawed frog)
Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 926)
REFERENCE
Croning,M.D.R., Ashurst,J.L., Taylor,R., Garrett,N. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (2004)
JOURNAL
COMMENT
Contact: Croning MDR
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS SEQUENCE ID: TTbA073h24.q1k7
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Nigel Garrett.
Seq primer: T7.
FEATURES
Location/Qualifiers
source
1..926
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TTbA073h24"
/dev_stage="tailbud (stage 28-30)"
/lab_host="Escherichia coli DH10B."
/clone_lib="XGC-tailbud"
/notes="Vector: pCSI07; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from 5' of poly A+ RNA from tailbud.
EcoRI-NotI cut cDNA was then ligated into pCSI07 with
EcoRI at the 5' end and NotI at the 3' end."

ORIGIN
Query Match 5.2%; Score 39; DB 7; Length 926;
Best Local Similarity 50.8%; Pred. No. 4.7;
Matches 93; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 8 AGATGACCGGAATGGTAAACGGAAGAAGGTCTGTGCATTTTATTTGCTGAGGCATTAAACAG 67
Db 186 AGAAGAACAGCGAGCTTATAGATTCATGACATGATACCAAGGCTGCCTGAGTGCTA 127
QY 68 AAACCTGCAGAGAAATATTGAAACGGCGCAGGATGTTTGGTGGGATCAGCCCAATTTA 127
Db 126 AGCTCTGAGTCCAAACATTTGCTCGATTGGATGTTTGTGGGTGGGGTGTATATA 67
QY 128 ACAGCAGGTTTCAGAGGATATATTACAGATTAATTGATGGCGGAAGCTCAATTTA 187
Db 66 GGGAGGGGTTTCAGAGGATTAATCAATNAATAAAGTAAATGTCAAAAAAATAAAAAA 7
QY 188 AAA 190
Db 6 AAA 4

RESULT 14
CNS017KX/c
LOCUS DEFINITION
CNS017KX 1101 bp DNA linear GSS 26-JUL-1999
Drosophila melanogaster genome survey sequence SP6 end of BAC
BACN37F10 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

QY 373 GATAACAAAGCTACCGAGCTGTTTCAGACAGAGATTGCAACATGCA 417
Db 118 GCAACAGATCCGATTCAGAGTCAGATCCAGAAAGTTGAGATGCA 74

ACCESSION AL108171 GI:5628475
VERSION AL108171.1
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
REFERENCE Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos of Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOBAC11.
FEATURES Location/Qualifiers
source
1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN37F10"
/clone_lib="DrosBAC"
/plasmid="pBelOBAC11"
/notes="end : SP6"

ORIGIN
Query Match 5.2%; Score 39; DB 9; Length 1101;
Best Local Similarity 21.9%; Pred. No. 4.9;
Matches 86; Conservative 135; Mismatches 165; Indels 7; Gaps 1;

QY 134 GTTTTCAGAGGATATATTACAGATTATTCGATGGCGGAAGCTCAATTTAAAGCG 193
Db 1090 DRTWTDKRRKKTWKTKADKKTRWTDKGGGGGKGGKGGKADTKRDWGAATR 1031
QY 194 TTTTCAGTTTACTTCGAGGCGATGAGCGGCTAATCTCTAGAGCGCTTGGAACTCGA 253
Db 1030 RARWATAATATATADTTWKRTTGTCTRTWKWARRAAKWRDWDWTDWMDTKWDK 971
QY 254 GAGGAAAGCTGAACGAAAGTAAGCAAGAGAGTATCACGAAACAGGAGATTTCGAGAA 313
Db 970 RWTWRWATGDDGAGRAKRRDRDRGRRRRRRTKRWWRRAAWTTATTTWWTGTTK 911
QY 314 GAGCCCTTGTGGCTCATGGCGGGATCCGAAAGCGATTTCATACATTTCTGATTTATAG 373
Db 910 GA-----WWTKTRKKGGAATAADWGRTDDWRWDKRAAAAGKGGKGAATWGAA 858
QY 374 ATAAACAAAGCTACAGCTGTTGAGACAGAGAGTTGAACTGATTTTTCAGAGCGCTC 433
Db 857 AAAAAAWDTKTATKTKRTKTGRAWAAAGTTRTTWDAATAWTKTAAAAAGRAAAAWAK 798
QY 434 ATTTTCGACATGCTGTTTGGACATGTCGTCGAAGGATAATCGGCGGTGGCGGGCGG 493
Db 797 DRGTGKGGKRRKRGTDGDKGTWKTKRTTWTATAAARRAARAAAKGWTKTDTGRRARA 738
QY 494 TCAGTTTGTATGTAAGAGAGTTCAGTTCAGGATA 526
Db 737 WDAWTTTTRKGAATRDGAAAAAATTTGGGAAA 705

RESULT 15
CNS00DLS
LOCUS DEFINITION
CNS00DLS 1021 bp DNA linear GSS 04-JUN-1999
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR27B02 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL072703

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2004, 12:45:13 ; Search time 84.1768 Seconds
(without alignments)
1069.666 Million cell updates/sec

Title: US-10-627-124-8
Perfect score: 1289
Sequence: 1 MNEMTGMVTRSVHPIA.....YLIVQEMAQLSEKRTISEG 251

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq_238ep04:.*
1: Geneseqp1980s:.*
2: Geneseqp1990s:.*
3: Geneseqp2000s:.*
4: Geneseqp2001s:.*
5: Geneseqp2002s:.*
6: Geneseqp2003as:.*
7: Geneseqp2003bs:.*
8: Geneseqp2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1289	100.0	251	8	ADJ62045
2	400.5	31.1	252	6	ABM70364
3	249	19.3	239	8	ADH02827
4	103.5	8.0	645	2	AAW29653
5	102.5	8.0	557	4	AAU20133
6	102.5	8.0	557	5	ABG91382
7	102.5	8.0	616	6	ABR41881
8	102.5	8.0	616	6	AAE36781
9	102.5	8.0	645	6	ABR41872
10	102.5	8.0	645	6	ABR41875
11	102.5	8.0	645	6	AAE36782
12	102.5	8.0	645	7	ADG31694
13	102.5	8.0	645	7	ADG31692
14	97	7.5	1554	7	ABO78068
15	95.5	7.4	2329	2	AAW25038
16	95.5	7.4	2637	4	AAU31618
17	95.5	7.4	3485	4	AAU31615
18	94.5	7.3	2986	4	AAU31620
19	94.5	7.3	3418	2	AAW23287
20	94.5	7.3	3418	2	AAW19211
21	94.5	7.3	3418	2	AAW04357
22	94.5	7.3	3418	2	AAW04356
23	94.5	7.3	3418	2	AAW04358
24	94.5	7.3	3418	2	AAW04354
25	94.5	7.3	3418	2	AAW04355

26	94.5	7.3	3418	3	AAW77819	AAW77819	BRCA2	pro
27	94.5	7.3	3418	8	ADJ32561	ADJ32561	Human	BRA
28	94.5	7.3	3418	8	ADK67819	ADK67819	Human	BRC
29	94.5	7.3	3418	8	ADL32565	ADL32565	Human	BRC
30	94.5	7.3	3423	4	ABG23417	ABG23417	Novel	hum
31	94.5	7.3	4741	4	AAU31612	AAU31612	Novel	hum
32	94.5	7.3	4743	4	AAU31619	AAU31619	Novel	hum
33	94.5	7.3	5292	4	AAU31617	AAU31617	Novel	hum
34	93.5	7.3	3418	7	ADJ68372	ADJ68372	Human	hea
35	93	7.2	480	4	ABB64489	ABB64489	Drosophil	
36	90	7.0	198	5	ABB98195	ABB98195	Arabidops	
37	90	7.0	198	5	AAW77963	AAW77963	Arabidops	
38	90	7.0	198	6	ABG75910	ABG75910	Arabidops	
39	90	7.0	198	7	ADP29071	ADP29071	Arabidops	
40	90	7.0	200	5	AAU93026	AAU93026	Arabidops	
41	90	7.0	200	7	ADC46691	ADC46691	Thalecres	
42	90	7.0	200	7	ADD30082	ADD30082	Plant yie	
43	90	7.0	200	8	ADI41695	ADI41695	Plant tra	
44	90	7.0	200	8	ADO02113	ADO02113	Thalecres	
45	90	7.0	534	2	AAW25031	AAW25031	Partial B	

ALIGNMENTS

RESULT 1
ADJ62045
ID ADJ62045 standard; protein; 251 AA.

XX AC ADJ62045;
XX DT 06-MAY-2004 (first entry)
XX DE Bacillus subtilis yvmC protein SEQ ID NO:8.
XX KW heterologous biological substance; Bacillus; yvmC; red pigment.
XX OS Bacillus subtilis.
XX PN WO2004011609-A2.
XX PD 05-FEB-2004.
XX PF 25-JUL-2003; 2003WO-US023398.
XX PR 26-JUL-2002; 2002US-0398853P.
XX PA (NOVO) NOVOZYMES BIOTECH INC.
XX Tang M, Sloma A, Sternberg D, Behr R;
XX WPI; 2004-143839/14.
XX N-PSDB; ADJ62044.

Producing a heterologous biological substance comprises cultivating pigment-deficient mutants of Bacillus cell that directs synthesis of the heterologous biological substance and has a modification of the *cypX* and *yvmC* genes.

Example 1; SEQ ID NO'8; 62pp; English.

The present invention describes a method for producing a heterologous biological substance comprising cultivating a mutant of a parent Bacillus cell in a medium suitable for the production of a heterologous biological substance, and recovering the heterologous biological substance from the cultivation medium, where the mutant cell comprises a first nucleic acid sequence directing synthesis of the heterologous biological substance and a second nucleic acid sequence comprising a modification of at least one of the genes *cypX* and *yvmC*, which are involved in the production of a red pigment. The mutant cell is also deficient in the production of the red pigment compared to the parent Bacillus cell when cultivated under the same conditions. Also described: (1) a mutant of a parent Bacillus cell, comprising a first nucleic acid sequence directing synthesis of a

Handwritten signature

CC heterologous biological substance and a second nucleic acid sequence
CC comprising a modification of at least one of the genes cypX and ymc,
CC which are involved in the production of a red pigment, wherein the mutant
CC cell is deficient in the production of the red pigment compared to the
CC parent Bacillus cell when cultivated under the same conditions; and (2)
CC obtaining a mutant of a parent Bacillus cell, comprising introducing into
CC the parent Bacillus cell a first nucleic acid sequence directing
CC synthesis of a heterologous biological substance and a second nucleic
CC acid sequence comprising a modification of at least one of the genes cypX
CC and ymc, which are involved in the production of a red pigment, and
CC identifying the mutant cell comprising the modified nucleic acid
CC sequence, where the mutant cell is deficient in the production of the red
CC pigment compared to the parent Bacillus cell when cultivated under the
CC same conditions. The methods and compositions of the present invention
CC are useful for producing biological substances, e.g. nucleic acid,
CC polyamide, polyamine, polyol, polypeptide, or polysaccharide, in pigment-
CC deficient mutants of Bacillus cells, and obtaining the pigment-deficient
CC mutant cells. The present sequence represents Bacillus subtilis ymcC,
CC which is used in the exemplification of the present invention.
XX
SQ Sequence 251 AA;

Query Match 100.0%; Score 1289; DB 8; Length 251;
Best Local Similarity 100.0%; Pred. No. 4.5e-143;
Matches 251; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNEMTGMVTRRSVHFTAEALTENCREFERRRHVLVGISPFNSRSEDYIYELIGWAKA 60
DB 1 MNEMTGMVTRRSVHFTAEALTENCREFERRRHVLVGISPFNSRSEDYIYELIGWAKA 60
QY 61 QFKSVSVLLAGHEAANLLEALGTPRGKAERKVRKESRNRFAERLVAHGGDPKAIHTF 120
DB 61 QFKSVSVLLAGHEAANLLEALGTPRGKAERKVRKESRNRFAERLVAHGGDPKAIHTF 120
QY 121 SDFIDNKAYQLLRQVEHAFPEQPHFRHACLDMSREAIIGRAGVSLMMEEVSEDMNLNLA 180
DB 121 SDFIDNKAYQLLRQVEHAFPEQPHFRHACLDMSREAIIGRAGVSLMMEEVSEDMNLNLA 180
QY 181 VEVVIAELPFFIGAPDILEVEETLLAYHRPWKLGEKISNHEFSICMRPNQGYLIVQEMAQ 240
DB 181 VEVVIAELPFFIGAPDILEVEETLLAYHRPWKLGEKISNHEFSICMRPNQGYLIVQEMAQ 240
QY 241 MLSEKRITSEG 251
DB 241 MLSEKRITSEG 251

RESULT 2
ABM70364
ID ABM70364 standard; protein; 252 AA.
XX
AC ABM70364;
DT 20-NOV-2003 (first entry)
XX
DE Photorhabdus luminescens protein sequence #3461.
XX
KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW detection; food; gene expression; plant; animal; microorganism; toxin;
KW antibiotic; biopesticide; virulence factor; disease model; plague;
KW whooping cough.
XX
OS Photorhabdus luminescens.
XX
FN WO200294867-A2.
XX
PD 28-NOV-2002.
XX
PF 07-FEB-2002; 2002WO-IB003040.
XX
PR 07-FEB-2001; 2001FR-00001659.
XX
PA (INSP) INST PASTEUR.

(CNRS) CNRS CENT NAT RECH SCI.
PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunet F, Danchin A;
PI Buchrieser C;
XX
DR WPI; 2003-148459/14.
XX
PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
PS Claim 2; SEQ ID NO 3461; 1205pp; French.
XX
CC The invention relates to the isolation of genes and their encoded
CC proteins from Photorhabdus luminescens. The isolated sequences are
CC sources of probes and primers for detecting the genome of P. luminescens
CC and related species; to study polymorphisms; for gene analysis and for
CC detection/amplification of the genes. Antibodies (Ab) raised against the
CC polypeptides encoded by the genes are used for detection/identification
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC carry a gene-containing vector are used to select compounds that
CC modulate, regulate, induce or inhibit expression of the genes in plants,
CC animals or microorganisms other than P. luminescens and are able to alter
CC response or sensitivity to toxins and antibiotics produced by P.
CC luminescens. Cells transformed to express the genes are useful for
CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luminescens proteins
XX
SQ Sequence 252 AA;
Query Match 31.1%; Score 400.5; DB 6; Length 252;
Best Local Similarity 37.6%; Pred. No. 3.5e-38;
Matches 86; Conservative 43; Mismatches 97; Indels 3; Gaps 2;
QY 13 SVHFIAEALTENCREFERRRHVLVGISPFNSRSEDYIYRLIGWAKAOKPSVLL-AG 71
DB 23 SPSFTVOGETSRCDQIIQGDHALIGISPFNSRFDYVVDLIQWSSHYPRQVDILLPCE 82
QY 72 HEANLLEALGTPRGKAERKVRKESRNRFAERLVAHGGDPKAIHT--FSPFIDNKAY 129
DB 83 REASRLLVASGIDNVKAIKTHREIRHRLNLDYVISTATLTKSKQIRVIOFSDFSLNHDY 142
QY 130 QLILRQVEHAFPEQPHFRHACLDMSREAIIGRAGVSLMMEEVSEDMNLNLAVEYVIAELP 189
DB 143 QSLKTQVENAFNESESFKSCLDMSFOAIKGRKLGITGQYFGQIDQLIVYKALPYIFAEIP 202
QY 190 FFIGAPDILEVEETLLAYHRPWKLGEKISNHEFSICMRPNQGYLIVQEM 238
DB 203 FYLNTPLRLGVKYSTLLYHRPWSIGKLFNGSGYPQIOWADKQSYGIVTQL 251
RESULT 3
ADH02827
ID ADH02827 standard; protein; 239 AA.
XX
AC ADH02827;
XX
DT 11-MAR-2004 (first entry)
XX
DE S. noursei AlbC SEQ ID NO:9.
XX
KW AlbC; diketopiperazine; antibacterial; fungicide; virucide;
KW immunosuppressant; cytostatic; cycloidiptide; COP; alpha;
KW beta-unsaturated diketopiperazine; DKP; aliboursine.
XX
OS Streptomyces noursei.
XX


```

XX Novel polypeptide useful for diagnosis, prognosis, prevention, and
PT treatment of immune, hyperproliferative, renal, respiratory, and
PT cardiovascular, reproductive, endocrine, gastrointestinal and
PT neurological disorders.
XX
PS Claim 11; Page 207-209; 296pp; English.
XX
XX The invention relates to novel isolated polypeptides (I) and
CC polynucleotides (II). (I) and (II) and antibody (III) to (I) are useful
CC for diagnosing a pathological condition or susceptibility to a
CC pathological condition in a subject and for preventing, treating or
CC ameliorating a medical condition. (I) is also useful for identifying a
CC binding partner to the polypeptide. (I), (II) and (III) are useful in
CC treating, preventing, diagnosing and/or prognosing immunodeficiencies
CC (e.g., B cell immunodeficiencies, severe combined immunodeficiencies),
CC autoimmune disorders (rheumatoid arthritis, multiple sclerosis, diabetes
CC mellitus), allergic reactions and conditions (e.g., asthma), inflammatory
CC conditions, graft-versus-host disease, blood-related disorders
CC (thrombosis, atherosclerosis), hyperproliferative disorders (e.g.,
CC cancer), renal disorders (e.g., acute glomerulonephritis), cardiovascular
CC disorders (e.g., arrhythmia), respiratory disorders (Goodpasture's
CC syndrome), neurological disorders (e.g., Alzheimer's disease, Parkinson's
CC disease), endocrine disorders (e.g., Addison's disease), reproductive
CC system disorders (e.g., endometriosis), infectious diseases (e.g., viral,
CC bacterial or fungal infections), and gastrointestinal disorders (e.g.,
CC Crohn's disease) (I) is useful to stimulate neuronal growth and treat,
CC prevent, and/or diagnose neuronal damage which occurs in certain neuronal
CC disorders or neuro-degenerative conditions, for stimulating keratinocyte
CC growth, to prevent hair loss, to modulate mammalian characteristics such
CC as body height, weight, hair colour, and to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors or other nutritional components. (I) is also useful
CC as a protein molecular weight marker and to raise antibodies. (II) is
CC useful for chromosome identification, radiation hybrid mapping, in gene
CC therapy, for identifying individuals from minute biological samples, as
CC additional DNA markers for restriction fragment length polymorphism
CC (RFLP), in forensic biology, as molecular weight markers on Southern
CC gels, as diagnostic probes for the presence of a specific mRNA in a
CC particular cell type, to raise anti-DNA antibodies using DNA immunisation
CC techniques, and as an antigen to elicit an immune response. (III) is
CC useful for immunophenotyping cell lines and biological samples and for
CC diagnosing and treating diseases, disorders or conditions. (III) is also
CC useful to assay protein levels in a biological sample. ABG91364-ABG91396
CC represent human amino acid sequences of the invention
XX
SQ Sequence 557 AA;
Query Match 8.0%; Score 102.5; DB 5; Length 557;
Best Local Similarity 26.3%; Pred. No. 0.017;
Matches 45; Conservative 24; Mismatches 71; Indels 31; Gaps 8;
QY 84 PRGKAERKVRKEVSRNRRFAERALVAHGDPKAIHTFSDFDINKAYQLLRQVEVHAFFEQ 143
Db 95 PEGATRKVIEKLAR-----FVAGGGPELEKVMEDYKDNPAFLDKNSREFL-- 144
QY 144 PFRHACLMSREAIIGRAGVSLMMEVESEDMNLNLAVEVYIAELPPFI--GAPDILEVE 201
Db 145 -YRKVKVAIRKEAQSQAASQKVSPPPE--DEEVKNLA-----EKLARFIADGGP---EVE 194
QY 202 ETLLAYHRPWKLGKISNHFESICMRPN-QGYLIVQEMAQMLSEKRITSEG 251
Db 195 TIALQNNR-----ENQAFSLYBPNQSGYKYRQKLEFRKAKASSTG 237
RESULT 7
ABR41881
ID ABR41881 standard; protein; 616 AA.
XX
AC ABR41881;
XX
DT 11-AUG-2003 (first entry)
XX
DE Human caspase-8 binding protein (p72) EST clone, THCS10568.

```

```

DE Predicted protein sequence of a human EST clone.
XX
XX Human; Cari; caspase-8 interacting protein; apoptosis; neuroprotective;
KW immunosuppressive; antidiabetic; hepatotropic; virucide;
KW antiinflammatory; antitumor; cytostatic; gene therapy;
XX expressed sequence tag; EST.
XX
OS Homo sapiens.
XX
XX WO2003020759-A2.
XX
XX 13-MAR-2003.
XX
XX 04-SEP-2002; 2002WO-IL000733.
XX
XX 04-SEP-2001; 2001IL-00145278.
PR 31-OCT-2001; 2001IL-00146251.
PR 06-JAN-2002; 2002IL-00147487.
XX
XX (YEDA ) YEDA RES & DEV CO LTD.
XX
XX Wallach D, Goncharov T, Kolumam G, Rajput A;
PI WPI; 2003-300863/29.
XX
XX Novel intracellular caspase-8 interacting polypeptide, designated as Cari
PT polypeptide, useful for treating inflammatory disease including multiple
PT sclerosis, autoimmune uveoretinitis and diabetes.
XX
XX Example 7; Page 116-119; 124pp; English.
XX
XX The present sequence is the predicted protein sequence of a human
CC expressed sequence tag (EST) clone obtained from a database screening
CC using tryptic peptides of an isolated 72.5 kDa protein identified as
CC Cari. The EST was used to screen a human gene database, and a cDNA (see
CC ACC48539) encoding Cari (see ABR41872) was subsequently obtained. Cari is
CC a novel intracellular caspase-8 interacting polypeptide. Cari
CC polypeptides, fragments, mutants, variants and derivatives, nucleic acids
CC encoding them, and anti-Cari antibodies can be used to treat an
CC inflammatory disease such as multiple sclerosis with primary
CC oligodendrogliaopathy, autoimmune uveoretinitis, diabetes, lupus,
CC autoimmune myocarditis I, hepatitis C virus (HCV) mediated chronic
CC hepatitis, chronic gastritis e.g. type A gastritis, mixed connective
CC tissue disease (MCTD), Crohn's disease and ulcerative colitis
XX
XX Sequence 616 AA;
Query Match 8.0%; Score 102.5; DB 6; Length 616;
Best Local Similarity 26.3%; Pred. No. 0.02;
Matches 45; Conservative 24; Mismatches 71; Indels 31; Gaps 8;
QY 84 PRGKAERKVRKEVSRNRRFAERALVAHGDPKAIHTFSDFDINKAYQLLRQVEVHAFFEQ 143
Db 154 PEGATRKVIEKLAR-----FVAGGGPELEKVMEDYKDNPAFLDKNSREFL-- 203
QY 144 PFRHACLMSREAIIGRAGVSLMMEVESEDMNLNLAVEVYIAELPPFI--GAPDILEVE 201
Db 204 -YRKVKVAIRKEAQSQAASQKVSPPPE--DEEVKNLA-----EKLARFIADGGP---EVE 253
QY 202 ETLLAYHRPWKLGKISNHFESICMRPN-QGYLIVQEMAQMLSEKRITSEG 251
Db 254 TIALQNNR-----ENQAFSLYBPNQSGYKYRQKLEFRKAKASSTG 296
RESULT 8
AAE36781
ID AAE36781 standard; protein; 616 AA.
XX
AC AAE36781;
XX
DT 07-AUG-2003 (first entry)
XX
DE Human caspase-8 binding protein (p72) EST clone, THCS10568.

```

XX Human; immunisation; caspase-8; enzyme linked immunosorbent assay; ELISA;
KW immunofluorescence; immunoelectron microscopy; caspase-8 binding protein;
KW p72 protein.
XX Homo sapiens.
XX WO2003020767-A2.
XX PD 13-MAR-2003.
XX PF 04-SEP-2002; 2002WO-IL000734.
XX PR 04-SEP-2001; 2001IL-00145279.
XX PA (YEDA) YEDA RES & DEV CO LTD.
XX PI Wallach D, Goncharov T, Kolumam G;
XX WPI; 2003-290170/28.
XX Novel antibody which co-immunoprecipitates caspase together with caspase-
PT bound protein, and releases caspase and bound protein from immune complex
PT upon elution, useful for purifying caspase-8 or caspase-bound protein.
XX Example 6; Page 68-70; 75pp; English.
XX The present invention relates to antibodies obtainable by immunisation of
CC an animal with a peptide from the C terminus end of the caspase-8 sub-1
CC unit and its fragments, capable of co-immunoprecipitating caspase (both
CC active caspase-8 and pro-caspase-8) together with a caspase-bound protein
CC and of releasing the caspase and the bound protein efficiently from the
CC immune complex upon elution. The antibodies of the invention are useful
CC for developing component of an enzyme linked immunosorbent assay (ELISA) and for
CC purifying caspase (preferably caspase-8) or caspase-bound protein. They
CC are useful to effectively co-precipitate and isolate a protein bound to
CC caspase-8. The antibodies can be employed histologically as in immuno-
CC fluorescence or immunoelectron microscopy for the in situ detection of
CC caspase-8. The present sequence is human caspase-8 binding protein (p72)
CC EST clone, THCS10568. This sequence is used in the exemplification of the
CC invention
XX
SQ Sequence 616 AA;
Query Match 8.0%; Score 102.5; DB 6; Length 616;
Best Local Similarity 26.3%; Pred. No. 0.02;
Matches 45; Conservative 24; Mismatches 71; Indels 31; Gaps 8;
QY 84 PRGAERKVKVSRNRRFAERLVAHGDPKAIHTFSDFDINKAYOLLQVEHAFPEQ 143
Db 154 PEGAETRKVIEKLAR-----FVARGGPELEKVMEDYKDNPAFAFLHDKNSREF-- 203
QY 144 PHERHACLDMSREAIIGRARGVSLMMEVESEDMNLNLAVEVIAELPFFI--GAPDILEVE 201
Db 204 -YRKVKVAEIRKGAQKSQAASQKVSPPPE-DEEVKNLA-----EKLARFIADGGP---EVE 253
QY 202 ETLLAYHRPWKLGKESINHEFSICMRPN-QGYLIVQEMQMLSEKRTITSEG 251
Db 254 TIALQNNR-----ENQAFSFLYBENSQGYKYRQKLEFRKAKASSTG 296
RESULT 9
ID ABR41872
XX ABR41872 standard; protein; 645 AA.
AC ABR41872;
XX 11-AUG-2003 (first entry)
XX Human caspase-8 interacting protein, Cari.
XX Human; Cari; caspase-8 interacting protein; apoptosis; neuroprotective;
KW immunosuppressive; antidiabetic; hepatotropic; virucide;

XX antiinflammatory; antiulcer; cytostatic; gene therapy.
XX Homo sapiens.
XX Key Location/Qualifiers
XX Domain 414..437
FT /note= "binds caspase-8, given in ABR41873"
FT Domain 422..437
FT /note= "binds caspase-8, given in ABR41874"
XX WO2003020759-A2.
XX PN 13-MAR-2003.
XX PD 04-SEP-2002; 2002WO-IL000733.
XX PF 04-SEP-2001; 2001IL-00145278.
XX PR 31-OCT-2001; 2001IL-00146251.
XX PR 06-JAN-2002; 2002IL-00147487.
XX PA (YEDA) YEDA RES & DEV CO LTD.
XX Wallach D, Goncharov T, Kolumam G, Rajput A;
XX WPI; 2003-300863/29.
XX N-PSDB; ACC48539.
XX Novel intracellular caspase-8 interacting polypeptide, designated as Cari
PT polypeptide, useful for treating inflammatory disease including multiple
PT sclerosis, autoimmune uveoretinitis and diabetes.
XX Claim 1; Page 120-122; 124pp; English.
XX The present sequence is the protein sequence of human Cari, a novel
CC intracellular caspase-8 interacting polypeptide. Cari activity is
CC required for induction of apoptosis through the tumour necrosis factor
CC receptor signalling pathway or by caspase-8. It binds to pro-caspase8 but
CC not to active caspase-8. Claimed pharmaceutical compositions comprising
CC Cari, its mutein, isoform, allelic variant, fragment, fusion protein or
CC derivative, a DNA sequence encoding these, an antisense or ribozyme
CC sequence, a vector comprising the DNA, or an anti-Cari antibody, are used
CC in the treatment of an inflammatory disease such as multiple sclerosis
CC with primary oligodendroglialopathy, autoimmune uveoretinitis, diabetes,
CC lupus, autoimmune myocarditis I, hepatitis C virus (HCV) mediated chronic
CC hepatitis, chronic gastritis e.g. Type A gastritis, mixed connective
CC tissue disease (MCTD), Crohn's disease and ulcerative colitis. A claimed
CC method for isolating polypeptides or factors involved in intracellular
CC signalling processes uses a Cari polypeptide, mutein, isoform, allelic
CC variant, fragment, fusion protein or derivative as the prey or bait in a
CC yeast two-hybrid procedure. A claimed method for screening for a peptide
CC or a small molecule antagonist to Cari comprises high-throughput
CC screening and selection of such molecules able to inhibit the interaction
CC of Cari to pro-caspase-8, or able to inhibit apoptosis enhanced by Cari
XX
SQ Sequence 645 AA;
Query Match 8.0%; Score 102.5; DB 6; Length 645;
Best Local Similarity 26.3%; Pred. No. 0.022;
Matches 45; Conservative 24; Mismatches 71; Indels 31; Gaps 8;
QY 84 PRGAERKVKVSRNRRFAERLVAHGDPKAIHTFSDFDINKAYOLLQVEHAFPEQ 143
Db 183 PEGAETRKVIEKLAR-----FVARGGPELEKVMEDYKDNPAFAFLHDKNSREF-- 232
QY 144 PHERHACLDMSREAIIGRARGVSLMMEVESEDMNLNLAVEVIAELPFFI--GAPDILEVE 201
Db 233 -YRKVKVAEIRKGAQKSQAASQKVSPPPE-DEEVKNLA-----EKLARFIADGGP---EVE 282
QY 202 ETLLAYHRPWKLGKESINHEFSICMRPN-QGYLIVQEMQMLSEKRTITSEG 251
Db 283 TIALQNNR-----ENQAFSFLYBENSQGYKYRQKLEFRKAKASSTG 325

QY 84 PRGAERKVRKEVSRNRRFAERLVAHGGDPKAIHTFSDFDINKAYOLLRQVEVHAFFEQ 143
Db 183 PEGAETRKVIEKLAR-----FVAGGGPELEKVMEDYKDNPAFAFLDKNSREFL-- 232
QY 144 PHERHACLDMSREALIGRAGVSLMMEEVSEDMNLAVEYVIAELPFFI--GAPDILEVE 201
Db 233 -YRKKAIRKEAQAQSKVSPPE--DEEVKNLA-----EKARFIADGGP---EVE 282
QY 202 ETLLAYHRPWKLGEKISNHEFSICMRPN-QGYLIVQEMAOQLSEKRTITSEG 251
Db 283 TIALQNNR-----ENQAFSLYEPNSQGYKYRQKLEEFKAKASSTG 325
RESULT 12
ADG31694
ID ADG31694 standard; protein; 645 AA.
AC ADG31694;
XX
XX 26-FEB-2004 (first entry)
DT Human protein with VEGF promoting activity SeqID40.
XX
XX human; vascularisation; vascular endothelial growth factor; VEGF;
KW tumour metastasis; psoriasis; chronic rheumatoid arthritis;
KW diabetic retinopathy; angiofibroma; macular degeneration;
KW angina pectoris; cerebral infarction; vascular dementia;
KW Burger's disease; antiinflammatory; cytostatic; antiarthritic;
KW antipsoriatic; antiarteriosclerotic; cardiant; vasotropic;
KW neuroprotective.
XX
OS Homo sapiens.
XX
XX WO2003083117-A2.
XX
XX 09-OCT-2003.
XX
XX 01-APR-2003; 2003WO-JP004163.
XX
XX 01-APR-2002; 2002JP-00098679.
PR 02-APR-2002; 2002US-0368977P.
XX
XX (ASAH) ASAH KASEI KK.
XX
XX Nagumo H, Matsuda A, Sugahara T;
XX
XX WPI; 2003-812544/76.
DR N-PSDB; ADG31693.
XX
XX New purified protein, useful for preparing a composition for treating or
PT preventing a disease associated with abnormal expression of VEGF or VEGF
PT receptor, e.g., solid tumor, inflammation, psoriasis or chronic
PT rheumatoid arthritis.
XX
XX Claim 1; SEQ ID NO 40; 405pp; English.
XX
XX This invention relates to novel proteins having vascularisation activity.
CC Specifically, it refers to proteins having an activity of promoting the
CC expression of vascular endothelial growth factor (VEGF) and/ or the VEGF
CC receptor. VEGF is a representative growth factor involved in normal
CC vascularisation in growing tissues, for example for fetal growth, tissue
CC repair and for the menstrual cycle and pregnancy. The present invention
CC describes these purified proteins, vectors and appropriate antibodies, as
CC well as a screening assay for identifying substances capable of
CC inhibiting or promoting VEGF (and VEGF receptor) expression. Furthermore,
CC it refers to compositions useful for the diagnosis, treatment and
CC prevention of diseases associated with abnormal VEGF/ VEGF receptor
CC expression, such as tumour metastasis, psoriasis, chronic rheumatoid
CC arthritis, diabetic retinopathy, angiofibroma, macular degeneration,
CC angina pectoris, cerebral infarction, vascular dementia or Burger's
CC disease. Accordingly, they exhibit antiinflammatory, cytostatic,
CC antiarthritic, antipsoriatic, antiarteriosclerotic, cardiant, vasotropic
CC and neuroprotective activities. This polypeptide sequence is a human

CC protein with VEGF promoting activity of the invention.
XX
SQ Sequence 645 AA;
Query Match 8.0%; Score 102.5; DB 7; Length 645;
Best Local Similarity 26.3%; Pred. No. 0.022;
Matches 45; Conservative 24; Mismatches 71; Indels 31; Gaps 8;
QY 84 PRGAERKVRKEVSRNRRFAERLVAHGGDPKAIHTFSDFDINKAYOLLRQVEVHAFFEQ 143
Db 183 PEGAETRKVIEKLAR-----FVAGGGPELEKVMEDYKDNPAFAFLDKNSREFL-- 232
QY 144 PHERHACLDMSREALIGRAGVSLMMEEVSEDMNLAVEYVIAELPFFI--GAPDILEVE 201
Db 233 -YRKKAIRKEAQAQSKVSPPE--DEEVKNLA-----EKARFIADGGP---EVE 282
QY 202 ETLLAYHRPWKLGEKISNHEFSICMRPN-QGYLIVQEMAOQLSEKRTITSEG 251
Db 283 TIALQNNR-----ENQAFSLYEPNSQGYKYRQKLEEFKAKASSTG 325
RESULT 13
ADG31692
ID ADG31692 standard; protein; 645 AA.
XX
AC ADG31692;
XX
XX 26-FEB-2004 (first entry)
DT Human protein with VEGF promoting activity SeqID38.
XX
XX human; vascularisation; vascular endothelial growth factor; VEGF;
KW tumour metastasis; psoriasis; chronic rheumatoid arthritis;
KW diabetic retinopathy; angiofibroma; macular degeneration;
KW angina pectoris; cerebral infarction; vascular dementia;
KW Burger's disease; antiinflammatory; cytostatic; antiarthritic;
KW antipsoriatic; antiarteriosclerotic; cardiant; vasotropic;
KW neuroprotective.
XX
OS Homo sapiens.
XX
XX WO2003083117-A2.
XX
XX 09-OCT-2003.
XX
XX 01-APR-2003; 2003WO-JP004163.
XX
XX 01-APR-2002; 2002JP-00098679.
PR 02-APR-2002; 2002US-0368977P.
XX
XX (ASAH) ASAH KASEI KK.
XX
XX Nagumo H, Matsuda A, Sugahara T;
XX
XX WPI; 2003-812544/76.
DR N-PSDB; ADG31691.
XX
XX New purified protein, useful for preparing a composition for treating or
PT preventing a disease associated with abnormal expression of VEGF or VEGF
PT receptor, e.g., solid tumor, inflammation, psoriasis or chronic
PT rheumatoid arthritis.
XX
XX Claim 1; SEQ ID NO 38; 405pp; English.
XX
XX This invention relates to novel proteins having vascularisation activity.
CC Specifically, it refers to proteins having an activity of promoting the
CC expression of vascular endothelial growth factor (VEGF) and/ or the VEGF
CC receptor. VEGF is a representative growth factor involved in normal
CC vascularisation in growing tissues, for example for fetal growth, tissue
CC repair and for the menstrual cycle and pregnancy. The present invention
CC describes these purified proteins, vectors and appropriate antibodies, as
CC well as a screening assay for identifying substances capable of
CC inhibiting or promoting VEGF (and VEGF receptor) expression. Furthermore,

CC it refers to compositions useful for the diagnosis, treatment and
CC prevention of diseases associated with abnormal VEGF/ VEGF receptor
CC expression, such as tumor metastasis, psoriasis, chronic rheumatoid
CC arthritis, diabetic retinopathy, angiofibroma, macular degeneration,
CC angina pectoris, cerebral infarction, vascular dementia or Buerger's
CC disease. Accordingly, they exhibit antiinflammatory, cytostatic,
CC antiarthritic, antipsoriatic, antiarteriosclerotic, cardiant, vasotropic
CC and neuroprotective activities. This polypeptide sequence is a human
CC protein with VEGF promoting activity of the invention.

XX
XX Sequence 645 AA;
Query Match 8.0%; Score 102.5; DB 7; Length 645;
Best Local Similarity 26.3%; Pred. No. 0.022;
Matches 45; Conservative 24; Mismatches 71; Indels 31; Gaps 8;
QY 84 PRGAERKVRKESVRRRFAERALVAHGGDPKAIHTFSDFDINKAYQLLRQVEHAFPEQ 143
DB 183 PEGAETRKVTEKLAR-----FVABGGPELEKVMEDYKDNPAFAFLHDKNSREFL-- 232
QY 144 PHFRHACLDMSRAITGRARGVSLMVEVSEDMNLNLAIVEVIAELPFFI--GAPDILEVE 201
DB 233 -YRKVKVAETIRKEAQKSAQKVSPPPE-DEEVKNLA-----EKLARFIADGGP---EVE 282
QY 202 ETLLAVHRPWKLGKISNHEFSICMRPN-OCYLIVQEMAOMLSEKRITSEG 251
DB 283 TIALQNRR-----ENQAFSLYEPNSQYKYRQKLEPRKAKASSTG 325

RESULT 14
ID ABO78068
XX ABO78068 standard; protein; 1554 AA.
AC ABO78068;
XX
XX 29-JUL-2004 (first entry)
DE Pseudomonas aeruginosa polypeptide #10243.
XX
XX Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.
XX Pseudomonas aeruginosa.
XX US651795-B1.
XX 22-APR-2003.
XX 18-FEB-1999; 99US-00252991.
XX 18-FEB-1998; 98US-0074788P.
XX 27-JUL-1998; 98US-0094190P.
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX WPI; 2003-615309/58.
XX N-PSDB; ABD11639.

XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
XX useful as molecular targets for diagnostics, prophylaxis and treatment of
XX pathological conditions resulting from bacterial infection.
XX
XX Disclosure; SEQ ID NO 26814; 455pp; English.
XX
XX The invention relates to Pseudomonas aeruginosa polypeptides and the
XX polynucleotides encoding them. The sequences are useful in diagnosis and
XX therapy of pathological conditions, as molecular targets for diagnostics,
XX prophylaxis and treatment of pathological conditions resulting from a
XX bacterial infection, for evaluating a compound, such as a polypeptide,
XX for the ability to bind a P. aeruginosa nucleic acid, as components of
XX effective antibacterial targets, as targets for antibacterial drugs,
XX including anti-P. aeruginosa drugs, as templates for recombinant

CC production of P. aeruginosa-derived peptides or polypeptides, as target
CC components for diagnosis and/or treatment of P. aeruginosa-caused
CC infection, and in detection of P. aeruginosa sequences or other sequences
CC of Pseudomonas species using biochip technology. Sequences ABO67826-
CC ABO84396 represent P. aeruginosa polypeptides of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html
XX
XX Sequence 1554 AA;

Query Match 7.5%; Score 97; DB 7; Length 1554;
Best Local Similarity 19.4%; Pred. No. 0.36;
Matches 64; Conservative 33; Mismatches 95; Indels 138; Gaps 14;
QY 1 MNEMTGMVTRSRVHFAIEALTENCREIFERRRHV-----LVGISPNSRFSEDIYRLI 55
DB 336 LGDATDAATDHADLHFVVAQAERALQSLERTAHIGLEDDVQGLLLFLAHVLED-VFQLA 394
QY 56 GWAKAQFKSVSVLLAGH-EAANILLEALGTGRGAERKVRK-----EVSRRNR--FAER-- 105
DB 395 GMGTROFHTLTELALTEQGFAGLL--LVGDHGHVLTGIRTVQAEGLDRDRRAGFLDLA 452
QY 106 ALVAHGGDPKAIHT-----
DB 453 VLVEHGADTTVVDADQHHVALAQGTVDQHGHHRAASLVQARLDHHTATARGRSQFOFEH 512
QY 120 -----FSDPID-----NKAYQLLRQVEHA-----FPEQPH 145
DB 513 FRLQQRFEQFIDASTELRRNRDEGSVAAPLFRDHVESQAVLDVVVRVGLGLVDLVRDH 572
QY 146 FRHA-----CLDMSREAIIGR-----ARGV-----SLMM 169
DB 573 DRHARRFVHLGFLGRHDAVVRGDHGDHIDIGRLGATGTHRGVAGGIGQGHHTLGF 632
QY 170 EEVSEDMNLNLAIVEVIAELPFFI GAPDILE 199
DB 633 DVVGTMDLGDATGLARGDL-----GATDVVE 658

RESULT 15
AAW25038
ID AAW25038 standard; protein; 2329 AA.
XX
XX AAW25038;
XX
XX 02-FEB-1998 (first entry)
DE Partial BRCA2 cancer susceptibility gene protein.
XX
XX BRCA2 cancer susceptibility gene; breast cancer; ovarian cancer;
XX gene therapy; prostate cancer; colorectal cancer; ocular melanoma;
XX leukaemia; human.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH Misc-difference 1267 /note= "Encoded by GA"
FT Misc-difference 1268 /note= "Encoded by AA"
FT
XX GB2307477-A.
XX
XX 28-MAY-1997.
XX
XX 25-NOV-1996; 96GB-00024453.
XX
XX 23-NOV-1995; 95GB-00023959.
XX 14-DEC-1995; 95GB-00025555.
XX 28-AUG-1996; 96GB-00017961.
XX
XX (CANC-) CANCER RES CAMPAIGN TECHNOLOGY.
PA

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2004, 13:27:20 ; Search time 21.4268 Seconds
(without alignments)
776.868 Million cell updates/sec

Title: US-10-627-124-8
Perfect score: 1289
Sequence: 1 MNEMTGMVTERRSVHFIAEA.....YLIVQEAQMLSEKRTISEG 251

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:
1: /cgn2_6/ptodata/1/iaa/5A COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	97	7.5	1554	4	US-09-252-991A-26814
2	95.5	7.4	2329	3	US-08-755-587-16
3	94.5	7.3	3418	2	US-08-639-501-2
4	94.5	7.3	3418	2	US-08-603-753D-4
5	94.5	7.3	3418	3	US-09-044-946-2
6	94.5	7.3	3418	3	US-08-755-587-44
7	94.5	7.3	3418	3	US-08-044-908-2
8	94.5	7.3	3418	3	US-08-089-753-4
9	94.5	7.3	3418	3	US-08-986-106-4
10	90	7.0	198	4	US-09-981-087A-25
11	90	7.0	198	4	US-08-978-382A-25
12	90	7.0	534	3	US-08-755-587-14
13	89.5	6.9	679	4	US-09-252-991A-28950
14	88	6.8	538	4	US-09-252-991A-22427
15	87	6.7	1021	4	US-09-252-991A-19205
16	84	6.5	819	3	US-09-651-656-15
17	84	6.5	819	3	US-09-650-855-15
18	84	6.5	880	4	US-09-252-991A-25046
19	82.5	6.4	394	2	US-08-646-590B-40
20	82.5	6.4	394	3	US-09-412-184-40
21	81.5	6.3	263	4	US-09-107-532A-6895
22	81.5	6.3	483	4	US-09-270-767-45923
23	81	6.3	420	4	US-09-583-110-5043
24	80.5	6.2	308	4	US-09-270-767-32667
25	80.5	6.2	308	4	US-09-270-767-47884
26	80	6.2	502	3	US-09-342-647-4
27	80	6.2	667	3	US-09-342-647-28

28	80	6.2	700	4	US-09-621-816B-2	Sequence 2, Appli
29	79.5	6.2	521	4	US-09-252-991A-25091	Sequence 25091, A
30	79.5	6.2	569	4	US-09-252-991A-29263	Sequence 29263, A
31	79.5	6.2	1206	4	US-09-245-928A-19	Sequence 19, Appl
32	79.5	6.2	1243	4	US-09-245-928A-17	Sequence 17, Appl
33	79.5	6.2	1257	3	US-08-947-823-5	Sequence 5, Appli
34	79.5	6.2	1257	4	US-09-245-928A-18	Sequence 18, Appl
35	78.5	6.1	249	4	US-09-543-681A-6793	Sequence 6793, Ap
36	78	6.1	487	1	US-08-452-722-7	Sequence 7, Appli
37	78	6.1	487	1	US-08-404-731A-7	Sequence 7, Appli
38	78	6.1	487	2	US-08-344-227-7	Sequence 7, Appli
39	78	6.1	487	2	US-08-503-226B-7	Sequence 7, Appli
40	78	6.1	487	3	US-08-721-458B-7	Sequence 7, Appli
41	77.5	6.0	289	4	US-09-395-674B-2	Sequence 2, Appli
42	77.5	6.0	363	3	US-08-984-618-13	Sequence 13, Appl
43	77.5	6.0	743	4	US-09-248-796A-15523	Sequence 15523, A
44	77	6.0	260	4	US-09-489-039A-12019	Sequence 12019, A
45	77	6.0	380	4	US-09-583-110-4774	Sequence 4774, Ap

ALIGNMENTS

RESULT 1
US-09-252-991A-26814
; Sequence 26814, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; SEQ ID NO 26814
; LENGTH: 1554
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26814

Query Match		7.5%	Score 97;	DB 4;	Length 1554;
Best Local Similarity		19.4%	Pred. No. 0.08;		
Matches		64;	Conservative	33;	Mismatches 95; Indels 138; Gaps 14;
Qy	1	MNEMTGMVTERRSVHFIAEALTE	NCREIFERRHV-----LVGISPFNSRPFSEYIVRLI	55	
Db	336	LGDATAATDHADLHFVVAQAERALQ	SLERTAHIGLEDDVQGLLFLAHVLE	VFQLA	394
Qy	56	GWAKAQKSVSVLLAGH-EAANLLEAL	GTPGKAEKVRK-----EVSRRNR--FAER--	105	
Db	395	GMGTROPHLTELALTEQGFAGLL--	LVGDHGLVLTGIRRTVQAEGLDRRRAGFLDRLA	452	
Qy	106	ALVAHGDDPKAIHT-----		119	
Db	453	VLVEHGADTTVDADQHHVALAQGT	VDQHGHHRAASLVQARLDHHAATARGRRSGFOFEH	512	
Qy	120	-----FSDPID-----	NKAYQLLRQVEHA-----	FFSQPH	145
Db	513	FRLQDRFEQFIDASTELRRNRDEGS	VAAFLFRDHVSSGOAVLDVVVRVGLGLVDLVRDH	572	
Qy	146	FRUA-----CLDMSREAIIGR-----	ARGV-----	SIMM	169
Db	573	DRHARRFVLHGLFGLRHDVAVGRD	HQDHDHIGRLGATGTHRGGRGVAGGIQEGHHATIGF	632	
Qy	170	EEVSEDMNLAVEYVIAELPFFIGAP	DILE 199		
Db	633	DVVGTDMLGDATGLARGDL-----	GATDVE 658		

RESULT 2
US-08-755-587-16
; Sequence 16, Application US/08755587
; Patent No. 6045997
; GENERAL INFORMATION:
; APPLICANT: Futreal, Phillip A
; APPLICANT: Wooster, Richard F
; APPLICANT: Ashworth, Alan
; APPLICANT: Stratton, Michael R
; TITLE OF INVENTION: Materials and methods relating to the
; TITLE OF INVENTION: identification and sequencing of the BRCA2 cancer
; TITLE OF INVENTION: susceptibility gene and uses thereof.
; NUMBER OF SEQUENCES: 222
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell Seltzer Park & Gibson
; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
; CITY: Raleigh
; STATE: NC
; COUNTRY: USA
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/755,587
; FILING DATE: 25-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9523959.6
; FILING DATE: 23-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9525555.0
; FILING DATE: 14-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9617961.9
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenneth D Sibley
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5405-135
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2329 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-755-587-16

Query Match 7.4%; Score 95.5; DB 3; Length 2329;
Best Local Similarity 18.4%; Pred. No. 0.23;
Matches 53; Conservative 57; Mismatches 109; Indels 69; Gaps 11;
QY 8 VTERRSVHFTAEALTENCRIFFERRHV-----LVGISPFNSRSEDIYVRLIG 56
DB 1828 VDKRNPCHVCNSEMEKTCSEFKLSNNLNVEGGSSNNHSHIKVSPYLSQFQDKQQLVLG 1887
QY 57 WAKAQFQSVSVLLAGHEAANLLEALGTPRGKAERKVRKVRNRFAERLVAHGGDPKA 116
DB 1888 TKVSLVENIHVL--GKEQA-----SPKNVKEIGKTFDSD---VPKTNIEV 1930
QY 117 IHTFS-----DFIDNKAYQLLRQVE-----HAFPEQPHRIACLDMSREAI 158
DB 1931 CSTYSKDSSENYFETEAVEIAKAFMEDDELTDKLPSSHATHSLFTCPNEEMVLNSR--- 1987
QY 159 IGRARGVSLAM--EEVSEDMN---LAVEVIAELPFFIGAPDILEVEETLLAYH--- 208
DB 1988 IGRARGVSLVGGPSIKRNLNLFEDRIENQEKSLKASKSTPDGTTIKDRRLFVHVSLE 2047
QY 209 -----RPWKLGEKISNHEFSICMRPNQGYLIIVQEMAQMLSEKRIITS 249
DB 2048 PITCVPRFTTKERQEIQNPNFT---APQGBFLSKSHLVHLETLKSSS 2092

RESULT 3
US-08-639-501-2
; Sequence 2, Application US/08639501
; Patent No. 5837492
; GENERAL INFORMATION:
; APPLICANT: Tavitgian, Sean V.
; APPLICANT: Kamb, Alexander
; APPLICANT: Simard, Jacques
; APPLICANT: Couch, Fergus
; APPLICANT: Remmens, Johanna
; APPLICANT: Weber, Barbara
; TITLE OF INVENTION: Chromosome 13-Linked Breast Cancer
; TITLE OF INVENTION: Susceptibility Gene
; NUMBER OF SEQUENCES: 124
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti
; STREET: 1201 New York Avenue N.W., Suite 1001
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 22204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/639,501
; FILING DATE: 29-APR-1996
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/585,391
; FILING DATE: 11-JAN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/576,559
; FILING DATE: 21-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/575,359
; FILING DATE: 20-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/573,779
; FILING DATE: 18-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-116802-04
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3418 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-639-501-2

Query Match 7.3%; Score 94.5; DB 2; Length 3418;
Best Local Similarity 18.4%; Pred. No. 0.53;
Matches 53; Conservative 57; Mismatches 109; Indels 69; Gaps 11;
QY 8 VTERRSVHFTAEALTENCRIFFERRHV-----LVGISPFNSRSEDIYVRLIG 56
DB 2109 VDKRNPCHVCNSEMEKTCSEFKLSNNLNVEGGSSNNHSHIKVSPYLSQFQDKQQLVLG 2168
QY 57 WAKAQFQSVSVLLAGHEAANLLEALGTPRGKAERKVRKVRNRFAERLVAHGGDPKA 116
DB 2169 TKVSLVENIHVL--GKEQA-----SPKNVKEIGKTFDSD---VPKTNIEV 2211
QY 117 IHTFS-----DFIDNKAYQLLRQVE-----HAFPEQPHRIACLDMSREAI 158
DB 2212 CSTYSKDSSENYFETEAVEIAKAFMEDDELTDKLPSSHATHSLFTCPNEEMVLNSR--- 2268

Qy 159 IGRARGVSLMM---EEVSEDMLN---LAVEVVIAPLFFGAPDILEVEETLLAYH--- 208
Db 2269 IGRGSGPLILVGPSPKRNLLNEFDRIENQKSLKASKSTPDGTTKDRRLFMHVSLE 2328
Qy 209 -----RPWKLGEKISNHEFSICMRPNQGYLIVQEMAQMLSEKRITS 249
Db 2329 PITCVPRTTKQEIQNPFT---APQBEFLSKSHLYEHLTLEKSSS 2373

RESULT 4
US-08-603-753D-4
; Sequence 4, Application US/08603753D
; Patent No. 5891857
; GENERAL INFORMATION:
; APPLICANT: HOLT, JEFFREY T.
; APPLICANT: JENSEN, ROY A.
; APPLICANT: PAGE, DAVID L.
; APPLICANT: KING, MARY-CLAIRE
; APPLICANT: SZABO, CSILLA I.
; APPLICANT: JETTON, THOMAS L.
; APPLICANT: ROBINSON-BENION, CHERYL L.
; APPLICANT: THOMPSON, MARILYN E.
; TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2
; TITLE OF INVENTION: PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON
; TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2 PROTEINS.
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARLES A. TAYLOR, JR.
; STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
; STREET: BOULEVARD
; CITY: DURHAM
; STATE: NORTH CAROLINA
; COUNTRY: USA
; ZIP: 27707
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800 KB storage
; COMPUTER: IBM PC/XT/AT compatible
; OPERATING SYSTEM: Windows 3.1
; SOFTWARE: WORD PERFECT 6.1 and ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/603,753D
; FILING DATE: 20 FEB 1996
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 08/373,799
; FILING DATE: 17 JAN 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: ARLES A. TAYLOR, JR.
; REGISTRATION NUMBER: 39,395
; REFERENCE/DOCKET NUMBER: 1242/2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919) 493-8000
; TELEFAX: (919) 419-0383
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3418
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: no
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens, sapiens
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE: adult
; TISSUE TYPE: female breast
; CELL TYPE: normal breast tissue
; CELL LINE: HMEC
; ORGANELLE: no
; FEATURE:

; NAME/KEY: BRCA2 protein
; LOCATION: 1 to 3418; Genbank locus HSU43746
; IDENTIFICATION METHOD:
; OTHER INFORMATION: BRCA2 protein has a negative
; OTHER INFORMATION: regulatory effect on growth of human mammary cells.
; PUBLICATION INFORMATION:
; AUTHORS: Wooster, R. et al.
; TITLE: Identification of the breast cancer
; TITLE: susceptibility gene BRCA2
; JOURNAL: Nature
; VOLUME: 379
; PAGES: 789-792
; DATE: 1995
; RELEVANT RESIDUES IN SEQ ID NO: 4: granin box
; RELEVANT RESIDUES IN SEQ ID NO: domain at amino acids 3334-3344
; US-08-603-753D-4

Query Match 7.3%; Score 94.5; DB 2; Length 3418;
Best Local Similarity 18.4%; Pred. No. 0.53; Mismatches 57; Indels 69; Gaps 11;
Matches 53; Conservative

Qy 8 VTERRSVHFAEALTENCREFERRHV-----LVGISPFNSRFSEDIYIRLIG 56
Db 2109 VDKRNPCHVSEMEKTSKEFKLSNNLNVGGSSNNHHSIKVSPYLSQFQDKQQLVLG 2168
Qy 57 WAKAQKSVSVLLAGHAANLLEALGTPRGKAEKVKVSRNRPAERALVAHGSDPKA 116
Db 2169 TKVSLVENIHVL--GKEQA-----SPKNVRMEIGTKTETFSN---VPVKTNI 2211
Qy 117 IHTFS---DFIDNKAYQLLRQVE-----HAFEPQHFHACLDMSREAI 158
Db 2212 CSTYSKDSNYFETEAVEIAKAFMEDDELDTSLPSHATSLFTCPENEEMVLSNSR--- 2268
Qy 159 IGRARGVSLMM---EEVSEDMLN---LAVEVVIAPLFFGAPDILEVEETLLAYH--- 208
Db 2269 IGRGSGPLILVGPSPKRNLLNEFDRIENQKSLKASKSTPDGTTKDRRLFMHVSLE 2328
Qy 209 -----RPWKLGEKISNHEFSICMRPNQGYLIVQEMAQMLSEKRITS 249
Db 2329 PITCVPRTTKQEIQNPFT---APQBEFLSKSHLYEHLTLEKSSS 2373

RESULT 5
US-09-044-946-2
; Sequence 2, Application US/09044946
; Patent No. 6033857
; GENERAL INFORMATION:
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Kamb, Alexander
; APPLICANT: Simard, Jacques
; APPLICANT: Couch, Fergus
; APPLICANT: Rommens, Johanna
; APPLICANT: Weber, Barbara
; TITLE OF INVENTION: Chromosome 13-Linked Breast Cancer
; TITLE OF INVENTION: Susceptibility Gene
; NUMBER OF SEQUENCES: 124
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti
; STREET: 1201 New York Avenue N.W., Suite 1001
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 22204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/044,946
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

```

; APPLICATION NUMBER: 08/639,501
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/576,559
; FILING DATE: 21-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/575,359
; FILING DATE: 20-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/573,779
; FILING DATE: 18-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-116802-04
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3418 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-044-946-2

```

Query Match	7.3%;	Score	94.5;	DB	3;	Length	3418;
Best Local Similarity	18.4%;	Pred. No.	0.53;				
Matches	53;	Conservative	57;	Mismatches	109;	Indels	69;
Gaps	11;						
QY	8	VTERRSVHFTAEALTCNCRIFERRRV-----LVGISPFNRSFSEDIYRLIG	56				
Db	2109	VDRKNPEHCNVSEMEKTCSEKFKLSNNLVEGGSENNHSIKVSPYLSFQDQKQQLVLG	2168				
QY	57	WAKAQPKSVSVILAGHEAANLLEALGTPRGKAERKVTKEVSRNRRFAERVALVAHGDPKA	116				
Db	2169	TKVSLVENIHVL---GREQA-----SPKNVKEIGKTEFSD---VPVKTNIEV	2211				
QY	117	IHTFS-----DFIDNKAYQLLRQVE-----HAFTEQPHFRHACLDMRSREAI	159				
Db	2212	CSTVSKDSSENYFETEAVEIAKAFMEDDELTDKLPSSHATHSLFTCPENEEMVLSNSR---	2268				
QY	159	IGRARGVSLMM---EEVSEDMNL---LAVEVVIATLPPFCADILEVEETLLAYH----	208				
Db	2269	IGRRGEPFLILVGEPSIKRNLLNEFDRIIEENQEKSLKASKSTPDGTTKDRLRFMHVSLSE	2328				
QY	209	-----RPWKLGEKISHNHFSCIMRPNQGYLIIVQEMAQMLSEKRITS	249				
Db	2329	PITCVPPRTTKEROEIQNPNT---APGQFLSKSHLYEHLTLEKSS	2373				

```

RESULT 6
US-08-755-587-44
; Sequence 44, Application US/08755587
; Patent No. 6045997
; GENERAL INFORMATION:
; APPLICANT: Futreal, Phillip A
; APPLICANT: Wooster, Richard F
; APPLICANT: Ashworth, Alan
; APPLICANT: Stratton, Michael R
; TITLE OF INVENTION: Materials and methods relating to the
; TITLE OF INVENTION: identification and sequencing of the BRCA2 cancer
; TITLE OF INVENTION: susceptibility gene and uses thereof.
; NUMBER OF SEQUENCES: 222
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell Seltzer Park & Gibson
; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
; CITY: Raleigh
; STATE: NC
; COUNTRY: USA
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
;
; CURRENT APPLICATION DATA:
;
; APPLICATION NUMBER: US/08/755,587
; FILING DATE: 25-NOV-1996
;
; PRIOR APPLICATION DATA:
;
; APPLICATION NUMBER: GB 9523959.6
; FILING DATE: 23-NOV-1995
;
; PRIOR APPLICATION DATA:
;
; APPLICATION NUMBER: GB 9525555.0
; FILING DATE: 14-DEC-1995
;
; PRIOR APPLICATION DATA:
;
; APPLICATION NUMBER: GB 9617961.9
; FILING DATE: 28-AUG-1996
;
; ATTORNEY/AGENT INFORMATION:
;
; NAME: Kenneth D Sibley
;
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5405-135
;
; INFORMATION FOR SEQ ID NO: 44:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 3418 amino acids
;
; TYPE: amino acid
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: protein
;
; US-08-755-587-44

```

Query Match	7.3%;	Score 94.5;	DB 3;	Length 3418;
Best Local Similarity	18.4%;	Pred. No. 0.53;		
Matches	53;	Conservative	57;	Mismatches 109; Indels 69; Gaps 11;
QY	8	VTRRSVHFPIARALTNCREIFERRHV-----LVGISPNRSPSEDIYIVRLIG	56	
Db	2109	VDKRNPEHCNVSEMEXTCKEFKLSNNLVEGGSSENNHSIKVSPYLSPQQDQQLVLG	2168	
QY	57	WAKAQPKSVSLLAGHAANLLALCTPRGKAERKYRKVERSRNRRAERALVAHGGDPKA	116	
Db	2169	TKVSLVENIHLV--GKEQA-----SPKNVMELCKTETPSD---VPVKTNIEV	2211	
QY	117	IHTFS----DFIDNKAYQLLRQVE-----HAFEPQHFRHACLDMSREAI	158	
Db	2212	CSTYSKDSENYPETEAVEITAKAFMEDDELTDSKLPGHATHSLFTCPENEEMVLSNR---	2268	
QY	159	IGRGAVGLMM-----EVSEDMLN---LAVEYVIAELPFFGIAGPDILEVEETLLAYH---	208	
Db	2269	IGRRGEPIILVCEPSIKKNLNLEFPRIENQEKSLKASKSTDGTGDIKDRPLFMHHVSL	2328	
QY	209	-----RPWKLGKISNHEFSCIMRPNQGYLIQVEMAOMLSKRITS	249	
Db	2329	PITCVPRFTTKRQEIQNPFT---APGOEFFLSKHLYEHLETLKSSS	2373	

RESULT 7
US-09-044-908-2
; Sequence 2, Application US/09044908
; Patent No. 6124104
; GENERAL INFORMATION:
; APPLICANT: Tavtighian, Sean V.
; APPLICANT: Kamb, Alexander
; APPLICANT: Simard, Jacques
; APPLICANT: Couch, Fergus
; APPLICANT: Rommens, Johanna
; APPLICANT: Weber, Barbara
; TITLE OF INVENTION: Chromosome 13-Linked Breast Cancer
; TITLE OF INVENTION: Susceptibility Gene
; NUMBER OF SEQUENCES: 124
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti
; STREET: 1201 New York Avenue N.W., Suite 1001
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 22204
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/044,908
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/639,501
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/576,559
FILING DATE: 21-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/575,359
FILING DATE: 20-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/573,779
FILING DATE: 18-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-116802-04
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3418 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-044-908-2

Query Match 7.3%; Score 94.5; DB 3; Length 3418;
Best Local Similarity 10.4%; Pred. No. 0.53; 109; Indels 69; Gaps 11;
Matches 53; Conservative 57; Mismatches 109; Indels 69; Gaps 11;

QY 8 VTERRSVHFAETALTECNREIFERRHV-----LVGIFPNSRFSSEDIYRLIG 56
DB 2109 VDKRNPCHCNSEMEKTCSEFKLSNNLNVEGSSNNHSIKVSPYLSQFQDQQLVLG 2168
QY 57 WAKAQKSVSVLLAGHAANLLEALGTGRGAERKVRKVRNRFAERLVAHGGDPKA 116
DB 2169 TKVSLVNIHVL--GKEQA-----SPKNVMEIGKTTETSD---VPVKTNIIV 2211
QY 117 IHTFS-----DPIDKAVQLLRQVE-----HAFPEQPHRACLDMSEAI 159
DB 2212 CSTYSKDSSENYFETEVEIAKAFMEDDELTDKSLPASHATSLFTCPENEEMVLSNR--- 2268
QY 159 IGRARGVSLMW---EEVSEMDLN---LAVEVIAELPFFIGAPDILEVEETLLAYH--- 208
DB 2269 IGRGRGPLILVGPSPKRNLLNEDFDRIENQEKSLKASKSTPDGTTKDRRLFMHVLSLE 2328
QY 209 -----RPWKLGGKSHNHSFICWRPNQGYLIIVQEMAQLSEKRITS 249
DB 2329 PITCVPTTKERQEIQNPNFT---AQQBELSKSHLYEHLTLEKSSS 2373

RESULT 8
US-09-099-753-4
Sequence 4, Application US/09099753
Patent No. 6149903
GENERAL INFORMATION:
APPLICANT: HOLT, JEFFREY T.
APPLICANT: JENSEN, ROY A.
APPLICANT: PAGE, DAVID L.
APPLICANT: KING, MARY-CLAIRE
APPLICANT: SZABO, CSILLA I.
APPLICANT: JETTON, THOMAS L.
APPLICANT: ROBINSON-BENION, CHERYL L.
APPLICANT: THOMPSON, MARILYN E.

TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2
TITLE OF INVENTION: PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON
TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2 PROTEINS.
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARLES A. TAYLOR, JR.
STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
CITY: DURHAM
STATE: NORTH CAROLINA
COUNTRY: USA
ZIP: 27707

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 800 kb storage
COMPUTER: IBM PC/XT/AT compatible
OPERATING SYSTEM: Windows 3.1
SOFTWARE: WORD PERFECT 6.1 and ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/099,753
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/603,753
FILING DATE: 20 FEB 1996
APPLICATION NUMBER: U.S. 08/373,799
FILING DATE: 17 JAN 1995
ATTORNEY/AGENT INFORMATION:
NAME: ARLES A. TAYLOR, JR.
REGISTRATION NUMBER: 39,395
REFERENCE/DOCKET NUMBER: 1242/2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 493-8000
TELEFAX: (919) 419-0383
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3418
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: no
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: Homo sapiens sapiens
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE: adult
TISSUE TYPE: female breast
CELL TYPE: normal breast tissue
CELL LINE: HMEC
ORGANELLE: no
FEATURE:
NAME/KEY: BRCA2 protein
LOCATION: 1 to 3418; Genbank locus HSU43746
IDENTIFICATION METHOD:
OTHER INFORMATION: BRCA2 protein has a negative
OTHER INFORMATION: regulatory effect on growth of human mammary cells.
PUBLICATION INFORMATION:
AUTHORS: Wooster, R. et al.
TITLE: Identification of the breast cancer
TITLE: susceptibility gene BRCA2
JOURNAL: Nature
VOLUME: 379
PAGES: 789-792
DATE: 1995
RELEVANT RESIDUES IN SEQ ID NO: 4: granin box
RELEVANT RESIDUES IN SEQ ID NO: domain at amino acids 3334-3344
US-09-099-753-4

Query Match 7.3%; Score 94.5; DB 3; Length 3418;
Best Local Similarity 10.4%; Pred. No. 0.53; 109; Indels 69; Gaps 11;
Matches 53; Conservative 57; Mismatches 109; Indels 69; Gaps 11;


```
Db 449 REAFHLVVEQLDNRLEIGRGEDVDHIAHTESRTGEIHV-----AGVLQA 496
QY 71 GHEAAN--LLEALGTRGKAERKVRKEVSNRRFAERALVAHGDDPKAIHTFSDFDINKA 128
Db 497 GQSPQQFALVDV-----AAVQVDHLQVGRGITEAVDGRHGDDDDRLTLEQRLGRQ 550
QY 129 YQLRQVEVHAFFEQPHFRHACLDMSREALIGRAR-GVSLMMEEVSEDMNLAVEYVIAE 187
Db 551 AHLLDMVD-----RGVLIDIG-----IGRDVGLRLVVVVGVNEVLHGIVREERLE 597
QY 188 LPPFIGAPDILEVEETLLAYHRPWKLGKESNHE-FSICMRPNQG 231
Db 598 LPVLGRQGLVRRQDQGRALH-----LGNHVGDASLARARHPQOG 638

RESULT 14
US-09-252-991A-22427
; Sequence 22427, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22427
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22427
```

```
Query Match 6.8%; Score 88; DB 4; Length 538;
Best Local Similarity 26.0%; Pred. No. 0.19;
Matches 44; Conservative 22; Mismatches 55; Indels 48; Gaps 8;

QY 60 AQPKSVSVLLAGHEAAN--LLE-----ALGTPRG-----KAERKVRKEVSNRRF----- 102
Db 92 AQPAYVARPVVGHQAGHGTVFERGAVEAQRGLLEEVFEQQQDVLAAVAQRQFQAHYF 151
QY 103 -----AERALVAHGG-----DPKAIH-----TFSDFDINKAYQLLRQVEHA- 139
Db 152 EPVVKVAAELAAALAHGGEVCLGRRDHPAVHRDLRVRAEAFQGVFLQHPQLDLQVDRHAL 211
QY 140 -FFEQPHFRHACLDMSREALIGRARGVSLMMEEVSEDMNLAVEYVIAE 187
Db 212 DLVEEQRAAVGMLDLADPLPLVGTGEGVRLVAE-----YLAVEQVLRQ 253
```

```
RESULT 15
US-09-252-991A-19205
; Sequence 19205, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19205
; LENGTH: 1021
```

```
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19205

Query Match 6.7%; Score 87; DB 4; Length 1021;
Best Local Similarity 25.7%; Pred. No. 0.65;
Matches 45; Conservative 31; Mismatches 67; Indels 32; Gaps 8;

QY 26 REIFERRRHV-----LVGISPFNSRSEDIYVRLIGWAKAQFKSVSVLLAGHEAANLL-E 79
Db 586 RDVEERLRLLHVDLTGLA--NRRLFRERLHEAQAQRQEGGLALLLLDLDRFKLLND 643
QY 80 ALGTPRGKAERKVRKEVSNRRRFAERA--LVAHGGDPKAI-----HTFSDFDINKAYQLLR 133
Db 644 SLG--HEVADQLLREMAQRLRTAPEANTLARLAGDEFAILLDDGGTGTAAALSRLAERLLV 701
QY 134 QVEVHAFFEQPHFRHACLDMSREALIGRARGVSLMMEEVSE-----DMLNLAVEY 183
Db 702 Q-----LRQPVSVLGHHELILGASIGISLFSQAREISVLMQANLAMQH 745
```

Search completed: December 10, 2004, 13:46:53
Job time : 23.4268 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2004, 13:45:11 ; Search time 73.4634 Seconds
(without alignments)
1220.359 Million cell updates/sec

Title: US-10-627-124-8
Perfect score: 1289
Sequence: 1 MNEMTGVTRRSHVHFAIA.....YLIVQMAQMLSEKRTISEG 251

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1585576 seqs, 357178320 residues

Total number of hits satisfying chosen parameters: 1585576

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first: 45 summaries

- Database : Published Applications AA:*
- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
 - 12: /cgn2_6/ptodata/1/pubpaa/US09D_PUBCOMB.pep.*
 - 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
 - 17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
 - 18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
 - 19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
 - 20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1289	100.0	251	15	US-10-627-124-8
2	102.5	8.0	557	9	US-09-764-856-60
3	102.5	8.0	557	11	US-09-764-856-60
4	102.5	8.0	557	14	US-10-102-627-60
5	100	7.8	1093	14	US-10-156-761-14825
6	96.5	7.5	1710	16	US-10-437-963-199352
7	94.5	7.3	3418	14	US-10-392-113-42
8	94.5	7.3	3418	15	US-10-634-574-1
9	93.5	7.3	495	16	US-10-437-963-114455
10	93.5	7.3	3418	16	US-10-408-765A-178
11	93	7.2	1156	14	US-10-369-493-21367
12	91.5	7.1	610	16	US-10-437-963-137309
13	90.5	7.0	545	16	US-10-437-963-148730

14	90	7.0	138	9	US-09-978-730-25	Sequence 25, Appl
15	90	7.0	138	9	US-09-978-729A-25	Sequence 25, Appl
16	90	7.0	138	9	US-09-981-087A-25	Sequence 25, Appl
17	90	7.0	138	9	US-09-978-382A-25	Sequence 25, Appl
18	90	7.0	138	10	US-09-978-740A-25	Sequence 25, Appl
19	90	7.0	200	10	US-09-934-455-128	Sequence 128, App
20	90	7.0	200	14	US-10-286-264-90	Sequence 90, Appl
21	90	7.0	200	14	US-10-225-066A-114	Sequence 114, App
22	90	7.0	200	15	US-10-374-780A-158	Sequence 158, App
23	90	7.0	200	15	US-10-412-699B-526	Sequence 526, App
24	90	7.0	1172	17	US-10-739-930-6385	Sequence 6385, Ap
25	88.5	6.9	284	15	US-10-282-122A-52738	Sequence 52738, A
26	87.5	6.8	724	14	US-10-369-493-13830	Sequence 13830, A
27	87	6.7	365	14	US-10-369-493-23225	Sequence 23225, A
28	87	6.7	977	16	US-10-437-963-124951	Sequence 124951, A
29	86.5	6.7	1024	16	US-10-437-963-124952	Sequence 124952, A
30	86.5	6.7	249	16	US-10-437-963-184335	Sequence 184335, A
31	86	6.7	179	14	US-10-365-742-74	Sequence 74, Appl
32	85.5	6.6	560	15	US-10-425-114-57637	Sequence 57637, A
33	85.5	6.6	776	17	US-10-425-115-228544	Sequence 228544, A
34	85.5	6.6	786	15	US-10-425-114-59864	Sequence 59864, A
35	85	6.6	716	14	US-10-369-493-774	Sequence 774, App
36	85	6.6	1828	16	US-10-437-963-123225	Sequence 123225, A
37	84.5	6.6	370	17	US-10-425-115-350281	Sequence 350281, A
38	84.5	6.6	679	16	US-10-437-963-185821	Sequence 185821, A
39	84.5	6.6	2033	16	US-10-408-765A-419	Sequence 419, App
40	84	6.5	419	14	US-10-369-493-16720	Sequence 16720, A
41	84	6.5	425	14	US-10-369-493-4922	Sequence 4922, Ap
42	84	6.5	425	14	US-10-369-493-7680	Sequence 7680, Ap
43	84	6.5	1013	16	US-10-437-963-199969	Sequence 199969, A
44	83.5	6.5	965	16	US-10-437-963-182421	Sequence 182421, A
45	83.5	6.5	1170	14	US-10-369-493-3006	Sequence 3006, Ap

ALIGNMENTS

RESULT 1
US-10-627-124-8
; Sequence 8, Application US/10627124
; Publication No. US20040096944A1
; GENERAL INFORMATION:
; APPLICANT: NOVOZYMES BIOTECH, INC.
; TITLE OF INVENTION: Methods For Producing Biological Substances In Pigment-Deficient
; FILE OF INVENTION: Mutants Of Bacillus Cells
; TITLE REFERENCE: 10302.200-WO
; CURRENT APPLICATION NUMBER: US/10/627,124
; CURRENT FILING DATE: 2003-07-25
; PRIOR APPLICATION NUMBER: US 60/398,853
; PRIOR FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 8
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-10-627-124-8

Query Match	100.0%	Score 1289;	DB 15;	Length 251;
Best Local Similarity	100.0%;	Pred. NO. 1e-128;		
Matches 251;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MNEMTGVTRRSHVHFAIALEALTCREIFERRRHVLGVISPFNSRFSSEDIYRLIGAWA	60	
Db	1	MNEMTGVTRRSHVHFAIALEALTCREIFERRRHVLGVISPFNSRFSSEDIYRLIGAWA	60	
Qy	61	QFKSVSVLLAGHAANLLEALGTGPKAERKVKRKEVSRNRRFAERLVAHGGPKAIHTF	120	
Db	61	QFKSVSVLLAGHAANLLEALGTGPKAERKVKRKEVSRNRRFAERLVAHGGPKAIHTF	120	
Qy	121	SDFIDNKAYQLLRQVEHFAFFEQPHFHAFLDMSREAIIGRARGVSLMVEVSDMLNLA	180	
Db	121	SDFIDNKAYQLLRQVEHFAFFEQPHFHAFLDMSREAIIGRARGVSLMVEVSDMLNLA	180	

Applicant

APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
PRIOR FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 14825
LENGTH: 1093
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-14825

Query Match 7.8%; Score 100; DB 14; Length 1093;
Best Local Similarity 26.5%; Pred. No. 0.49;
Matches 36; Conservative 15; Mismatches 47; Indels 38; Gaps 4;

QY 19 EALTENC-----REIFERRHVLGISPSPNSRFSYRLLIGWAKA 61
DB 939 EALRHCDRAGVPYTTADGRGVVLEMYERLVEERTGLTFYKDFPTD----- 986
QY 62 FKSVSLLAGHEA-ANLE-----ALGTPGKAEKVKVSRNRRFAERALVAHGGDPK 115
DB 987 ---VSLTROHARDPRLAERWDLVAFGTGLTAYSELTDPEQRRLTAQSLAANGDPE 1043
QY 116 AIHTFSDFDINKAYQL 131
DB 1044 AMELDEDFDLAEYAM 1059

RESULT 6
US-10-437-963-199352
Sequence 199352, Application US/10437963
Publication No. US20040123343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazov, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 199352
LENGTH: 1710
TYPE: PRT
ORGANISM: Oryza sativa
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_94925C.1.pap
US-10-437-963-199352

Query Match 7.5%; Score 96.5; DB 16; Length 1710;
Best Local Similarity 24.0%; Pred. No. 2.2;
Matches 41; Conservative 26; Mismatches 41; Indels 63; Gaps 8;

QY 54 LIGWAKAQFKSVSVLLAGHEAANLEALGTPGKAEKVKVSRNRRFAERALVAHGGD 113
DB 889 LQGSVKVKFNVS-----RAIEKARDUA-----TLVSHNAD 920
QY 114 PKAHTFSDFDINKAYQ-----LLR-----QVEHAFPEQPHFRHACLD 153

DB 921 YSEIRSSDVLNELLRYEEMLWLQSRISWLKEVKSTHEMEKIETDYFKOMFTADPCIDQ 980
QY 154 SREAIIGRAGVSIIMEEVSEDM-LNLAVEV-----IAELPFFIG--APDI 197
DB 981 SRVS-----RLFQKVSPFEMNVLDCKDFTQEQIADALFQIGPIKAPGL 1023

RESULT 7
US-10-392-113-42
Sequence 42, Application US/10392113
Publication No. US20030224993A1
GENERAL INFORMATION:
APPLICANT: Land, Hartmut
APPLICANT: Deleu, Laurent
TITLE OF INVENTION: COMPOSITIONS THAT INHIBIT PROLIFERATION
TITLE OF INVENTION: OF CANCER CELLS
FILE REFERENCE: 21108.0005U3
CURRENT APPLICATION NUMBER: US/10/392,113
CURRENT FILING DATE: 2003-03-17
PRIOR APPLICATION NUMBER: 60/365,078
PRIOR FILING DATE: 2002-03-15
PRIOR APPLICATION NUMBER: PCT/US01/32127
PRIOR FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 60/239,705
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 45
SOFTWARE: Fast-Seq for Windows Version 4.0
SEQ ID NO 42
LENGTH: 3418
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence./Note =
OTHER INFORMATION: Synthetic Construct
US-10-392-113-42

Query Match 7.3%; Score 94.5; DB 14; Length 3418;
Best Local Similarity 18.4%; Pred. No. 9.6;
Matches 53; Conservative 57; Mismatches 109; Indels 69; Gaps 11;

QY 8 VTERRSVHFTAEALTENCRIEIPERRHV-----LVGISPFNRSFSEYIRLIG 56
DB 2109 VDKRNPCHVCNSEMECTSKFELSNLNVGSGSENHHSIKVSPYLSQFQDQQLVLG 2168
QY 57 WAKAQFKSVSVLLAGHEAANLEALGTPGKAEKVKVSRNRRFAERALVAHGGDPKA 116
DB 2169 TKVSLVENIHVL--GKEQA-----SPKNVKMEIGKTTFTSD---VPVKTNIEV 2211
QY 117 IHTFS-----DFIDNKAYQLLRQVE-----HAFPEQPHFRHACLDMSREAI 158
DB 2212 CSTYKSDSENYFTEAVEIAKAFMEDDELTDKLPSHATHSLFTCPNEEWLSNSR--- 2268
QY 159 IGRAGVSLMM-----EVSSEDMLN---LAVEYIAELPFFIGAPDILEVEETLAYH---- 208
DB 2269 IGKRRGEPLILVGEPSIKRNLNLEFDRIIENQEKSLKASKSTPDGTIKDRRLFMHVSLE 2328
QY 209 -----RPMKLGKISNHEFSICMRPNQCYLIVQEAQMLSEKRITS 249
DB 2329 PITCVPRFTTKERQEIQNPNFT---APQBFSLKSHLYEHLTLEKSSS 2373

RESULT 8
US-10-634-574-1
Sequence 1, Application US/10634574
Publication No. US20040072268A1
GENERAL INFORMATION:
APPLICANT: The Wistar Institute
APPLICANT: Ramin Shiekhatar
TITLE OF INVENTION: METHODS FOR REGULATING BRCA1-BRCA2-CONTAINING COMPLEX ACTIVITY
FILE REFERENCE: WSTR-0014B
CURRENT APPLICATION NUMBER: US/10/634,574
CURRENT FILING DATE: 2003-08-05

	Prior Application Number:	US 60/401,433
	Prior Filing Date:	2002-08-05
	Prior Application Number:	US 60/449,950
	Prior Filing Date:	2003-02-24
	Number of Seq ID NOS:	26
	Software:	PatentIn version 3.1
	Seq ID NO 1	
	Length:	3418
	Type:	PRT
	Organism:	Homo sapiens
	US-10-634-574-1	
	Query Match	7.3%; Score 94.5; DB 15; Length 3418;
	Best Local Similarity	18.4%; Pred. No. 9.6;
	Matches	53; Conservative 57; Mismatches 109; Indels 69; Gaps 11;
QY	8 VTERSVPFIAETALNCREIFERRHV-----LVGISPFNSRSEDVIYLIG 56	
DB	2109 VKRNPEHCVMSEKTSKEFKLSNNLNVEGSSNHNHSIKVSPYLSQFOODKQLVLG 2168	
QY	57 WAKAOKFSVVLLAGHEANLLEALGTPRGKAERKVRSNRFAERALVAHGGDPKA 116	
DB	2169 TKSVLENTHVL--GKEQA-----SPKNVKMEIKGTETFS---VPVKTNIIEV 2211	
QY	117 IHTFS----DFIDNKAYOLLROEVE-----HAFFEQPHFRACLDMSREAI 158	
DB	2212 CSTYSKDSENYPETEAVEITAKAFMEDDELTDKLPSSHATHSLFTCPENEEVWLSNR--- 2268	
QY	159 IGRARGVSLMM---EVSDMLN---LAVEYVIALPFIAPDILEVEETLLAYH--- 208	
DB	2269 IGKRGEPLILVGEPISIKRLNEFDRIENOESKLKASKSTPDGTIKDRFLFMHHVSLE 2328	
QY	209 -----RPWLGEKISNHFSICMRPNQGYYLIIVQMAMLSRKITS 249	
DB	2329 PITCVPRITTKERQEIQNPNT---APQQEFLSKSHLYEHULTLEKSSS 2373	
	RESULT 9	
	US-10-437-963-114455	
	; Sequence 114455, Application US/10437963	
	; Publication No. US20040123343A1	
	; GENERAL INFORMATION:	
	; APPLICANT: La Rosa, Thomas J.	
	; APPLICANT: Kovalic, David K.	
	; APPLICANT: Zhou, Yihua	
	; APPLICANT: Cao, Yongwei	
	; APPLICANT: Wu, Wei	
	; APPLICANT: Boukharov, Andrey A.	
	; APPLICANT: Barbazuk, Brad	
	; APPLICANT: Li, Ping	
	; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With	
	; FILE REFERENCE: 38-21(53221)B	
	; CURRENT APPLICATION NUMBER: US/10/437,963	
	; CURRENT FILING DATE: 2003-05-14	
	; NUMBER OF SEQ ID NOS: 204966	
	; SEQ ID NO 114455	
	; LENGTH: 495	
	; TYPE: PRT	
	; ORGANISM: Oryza sativa	
	; FEATURE:	
	; OTHER INFORMATION: Clone ID: PAT_MRT4530_18144C.1.pep	
	US-10-437-963-114455	
	Query Match	7.3%; Score 93.5; DB 16; Length 495;
	Best Local Similarity	22.7%; Pred. No. 0.79;
	Matches	57; Conservative 32; Mismatches 97; Indels 65; Gaps 12;
QY	49 DTYIRLGHWAKAQFSVLLAGHEANLLEALGTPRG-KAEKKVRKE-VSRNRRFAER 105	
DB	115 DTINONCHWSLLPLLALKIIFAG--LANIYGPSVTFGSRTDLANDSGVMSLRLAAS 172	
QY	106 ALVHGDDPKAHITFSDTINKAYOLLROEVEHAFFEOPHFRAACL-----DNMREAII 159	

GENERAL INFORMATION:
; APPLICANT: Yanofsky, Martin F.
; APPLICANT: Liljgren, Sarah
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Selective Control of Lignin Biosynthesis in Transgenic
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 19452A-000920US
; CURRENT APPLICATION NUMBER: US/09/978,730
; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 60/090,649
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: US 09/339,998
; PRIOR FILING DATE: 1999-06-25
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
; LENGTH: 198
; TYPE: PRT
; ORGANISM: Arabidopsis sp.
US-09-978-730-25

Query Match 7.0%; Score 90; DB 9; Length 198;
Best Local Similarity 29.3%; Pred. No. 0.51;
Matches 39; Conservative 22; Mismatches 46; Indels 26; Gaps 8;

QY 97 SRNRFAERALVAHGGDPKAIHTFSDFDINKAYQLLRQVEHAFFEQPHERHACLDSMSRE 156
DB 20 SRNHSPPKRSWM-----EPQPHLLMDW--NKANDLLTQE-HAALNDPH--HMLDPPPE 70
QY 157 AIIGRARGVSLMMEVSESDMLNL-AVEYVIAEL-----PFFIGAPD-----ILEVEETL 204
DB 71 TLHLDED-----EYDEDEDAMKEMQYMIAMQPVVDIDPATVPKPNRRNRISDDPQTV 125
QY 205 LAYHRPWKLGEKI 217
DB 126 VARRRRRISEKI 138

Search completed: December 10, 2004, 14:02:31
Job time : 75.4634 secs

RESULT 15
US-09-978-729A-25
; Sequence 25, Application US/09978729A
; Patent No. US20020178465A1
; GENERAL INFORMATION:
; APPLICANT: Yanofsky, Martin F.
; APPLICANT: Liljgren, Sarah
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Selective Control of Lignin Biosynthesis in Transgenic
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 19452A-000950US
; CURRENT APPLICATION NUMBER: US/09/978,729A
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: US 60/090,649
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: US 09/339,998
; PRIOR FILING DATE: 1999-06-25
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
; LENGTH: 198
; TYPE: PRT
; ORGANISM: Arabidopsis sp.
US-09-978-729A-25

Query Match 7.0%; Score 90; DB 9; Length 198;
Best Local Similarity 29.3%; Pred. No. 0.51;
Matches 39; Conservative 22; Mismatches 46; Indels 26; Gaps 8;

QY 97 SRNRFAERALVAHGGDPKAIHTFSDFDINKAYQLLRQVEHAFFEQPHERHACLDSMSRE 156
DB 20 SRNHSPPKRSWM-----EPQPHLLMDW--NKANDLLTQE-HAALNDPH--HMLDPPPE 70
QY 157 AIIGRARGVSLMMEVSESDMLNL-AVEYVIAEL-----PFFIGAPD-----ILEVEETL 204
DB 71 TLHLDED-----EYDEDEDAMKEMQYMIAMQPVVDIDPATVPKPNRRNRISDDPQTV 125
QY 205 LAYHRPWKLGEKI 217
DB 126 VARRRRRISEKI 138

Query Match 7.0%; Score 90; DB 9; Length 198;
Best Local Similarity 29.3%; Pred. No. 0.51;
Matches 39; Conservative 22; Mismatches 46; Indels 26; Gaps 8;

QY 97 SRNRFAERALVAHGGDPKAIHTFSDFDINKAYQLLRQVEHAFFEQPHERHACLDSMSRE 156
DB 20 SRNHSPPKRSWM-----EPQPHLLMDW--NKANDLLTQE-HAALNDPH--HMLDPPPE 70
QY 157 AIIGRARGVSLMMEVSESDMLNL-AVEYVIAEL-----PFFIGAPD-----ILEVEETL 204
DB 71 TLHLDED-----EYDEDEDAMKEMQYMIAMQPVVDIDPATVPKPNRRNRISDDPQTV 125
QY 205 LAYHRPWKLGEKI 217
DB 126 VARRRRRISEKI 138

95/0401
 25 (bases 30303 to 37035)
 REFERENCE
 ROBINSON,C., RIVOLTA,C., KARAMATA,D. and MOIR,A.
 AUTHORS
 The product of the ywoC (gefF) gene of *Bacillus subtilis* is
 TITLE
 required for spore germination
 JOURNAL
 Microbiology 144 (Pt 11), 3105-3109 (1998)
 MEDLINE
 99061203
 PUBMED
 9846746
 3 (bases 1 to 47739)
 REFERENCE
 LAZAREVIC,V., SOLDÓ,B., RIVOLTA,C., REYNOLDS,S., MAUEL,C. and
 AUTHORS
 Karamata,D.
 TITLE
 Nucleotide sequence of the 300-304 chromosomal segment of *Bacillus subtilis*
 JOURNAL
 Unpublished
 REFERENCE
 4 (bases 1 to 47739)
 AUTHORS
 Lazarevic, V.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (05-AUG-1997) IGBM, Cesar-Roux 19, Lausanne 1005,
 Switzerland
 FEATURES
 Location/Qualifiers
 source
 1. .47739
 /organism="Bacillus subtilis"
 /mol_type="genomic DNA"
 /db_xref="taxon:1423"
 /map="300-304 degrees"

```

gene
RBS
CDS

71. .1184
/gene="prfB"
71. .77
/gene="prfB"
join(83. .154,156. .1184)
/gene="prfB"
/notes="at low concentrations of RF-2 nucleotides 155. .158
(UGAC) are translated as Asp, Pcl, H.J. Nucleic Acids Res.
20. 4423-4428, 1992"
/codon_start=1
/transl_table=11
/product="putative peptide chain release factor RF-2"
/protein_id="AAC67303.1"
/db_xref="GI:2618874"
/translation="MELSEIRAELENWASRLADFRGSLDLSEKEARIELDEQADPE
FNDQQAQIVINEAGLKYDYNYSKLNESHEELQMTDHLKKEPDITDLQLEKEBL
KSTLRFNFEPLQLLSYDKNNAILELHPGAGGTESQDQWGSMLLMYTRWGERGF
KVTLDYLPQDEAGIKSVTLIKGHNAYGYLKAEGVHRLVRI SPFDSSGRRTSPVS
CEVMPFDEIDIDIRTEIDKVDTRYASGAGGOHVNTDSAVRI THLPNTNVVTCOTE
RSQIKRERAMKAKLYQRRIEEOQAEDELIRGEQKEIGWSQIRSVYFHPYSVK
DHRNTMGVQVMDGDIDTFIDAYLRSLKLS"
143. .149
/gene="prfB"
1338. .2195
/gene="yvjA"
1338. .1342
/gene="yvjA"
1350. .2195
/gene="yvjA"
/notes="similar to thermophilic bacterium PS-3 hypothetical
protein: PIR Accession Number S43727 and to Bacillus
subtilis Yqfu, YpJc and YitB proteins"
/codon_start=1
/transl_table=11
/product="yvjA"
/protein_id="AAC67260.1"
/db_xref="GI:2618831"
/translation="MDVRNKTLMIRDYVYLIGAAITAVSNVFLPNKIAAGVSG
ISTLOSYPEAAYVOMIINIPFIAGVILLGKFGKLTAGSVFLPVFLFRDIQOP
ATHHELLAIFGVGIGIGIVYLGSGTGTGALAAQI IHKYSLSLKGKCLAIIDGM
IVVTAMVFNIEQLYAMLGVVYSSKTI DVVQGFNRSKMALIITKQEQAKBAVLQK
IDRGVTKISAGGVTDDDRPILCVVGQTETFKLKQIVKQIDESA FVI VADASEVLGE
GFKRA"
2223. .2582
/gene="cccB"
2223. .2229
/gene="cccB"
2244. .2582
/gene="cccB"
/notes="putative lipoprotein"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="cytochrome c-551"
/protein_id="AAC67261.1"
/db_xref="GI:2618832"
/translation="MSKSIIMIGFALSVLAAACSNDAKEBKTDTGSKTEATASEG
EELYQSCVCGCHGKOLEGVSGPNLQEVGKYDHEKHIESI IKNGRGNMPLGLVDNEAA
VIAKWLSKK"
2806. .3502
/gene="ftsE"
2806. .2813
/gene="ftsE"
2816. .3502
/gene="ftsE"
/codon_start=1
/transl_table=11
/product="cell division ATP-binding protein"
/protein_id="AAC67262.1"
/db_xref="GI:2618833"
/translation="MIEMKEVYKAYPNGVKALNGISVTHHPGFVYVVGSGAGKSTFF
IKMIYREKPTKGLIINHDLATIKEI PFVRRKIGVVQDFKLLPKLTVFNVAF
ALVEIGQPSVKKRVLEVDLQVKHAKQFPDQLSGGQQRVSARSIVNNPDVVI

ADEPTGNLDPDTSWEVMKTLBENNRRGTTVMATHNKEIVNTMKRVAIEDGIIVRD
BSRGEYSYD"
3476. .4385
/gene="ftsX"
3476. .3484
/gene="ftsX"
3495. .4385
/gene="ftsX"
/codon_start=1
/transl_table=11
/product="cell division protein"
/protein_id="AAC67264.1"
/db_xref="GI:2618835"
/translation="MIKILGRHLRESFKSLGRNTWMTFASISAVTVTLILVGVFLVIM
LNLNWTNAEKQVEIKVLIDITADQAKQKQNDIKELKGIQSVTFSSKEKELDQV
DSFGSGKSLTMDQENLADAFVKTIDPHDTPNVAKKIEKMDHYIKVTYKBEVS
LFWVGVSNIGIALIIGLVFTAMFLISNTIKITIFARRKEIEMIKLVGATNFWIRP
FLEGILGVFGSVIPIALVLSYQYVIGVWVPKVQGSFVSLLPYTPFPVFQVSLVIA
IGAVIGWGLSTSIKRLRV"
4511. .5967
/gene="yvjB"
4511. .4516
/gene="yvjB"
4525. .5967
/gene="yvjB"
/codon_start=1
/transl_table=11
/product="putative protease"
/protein_id="AAC67263.1"
/db_xref="GI:2618834"
/translation="MNOKIMAVIAGSMFLGGAGVAGINLLEMDXPQTAAPATAQA
DSERDAMDKIEKAYELISNEYVEKYDREKLEGAIQGMLSTLNDPSVYMDKQTAQ
FSDSLDSSEGGIAGVGMEDKIIIVSPFKSPAEKAGLKPNDIEIISINGESMAGDL
NHAULKIRGKSGSVSMKIQRPCTKKQLSFRIKRAEIPLETIVFASKKVQGVHGTIA
ISTFSEHTAEDFAKALRELEKEIEGLVDVRNPGGYLQSVBEEILKHFVTKDQPYIQ
GTVOQAVPMGDGNSIKITLYKWLTPNGNIHKGI EPTTAIKQPDVFSAGPLQKEPL
IAERNGDKKRYSTLHKKAYPNVITDKGSASAEILAGALKAEAGHYDVVGDTSEK
IATGNGEDYKHAQVLLKGLSFDGREDGYFSKDMKAVAFQDNKLNKTGVIDTRTA
ETLNOQIEKSKSDERNDLQQLTALKSLFN"
6773. .7977
/gene="yvjD"
6773. .6777
/gene="yvjD"
6784. .7977
/gene="yvjD"
/notes="transmembrane protein"
/codon_start=1
/transl_table=11
/product="yvjD"
/protein_id="AAC67265.1"
/db_xref="GI:2618836"
/translation="MSVQNGIELLSAGLFFLHPLFWFFIITLARGYVRIKERTKF
HTRIADYDDLKFTYTKGLIPGLLSVILGGIGISIPGLLAIIVITAAAFTRAN
WMSAAYIVSVMLIGFLQIYQAEPLERFPQGFVAVWPAVAVLGLLITTEGAVYR
SAHRTSPALVSSRGLPIQOOLANRWLLPLFLVPGNGLESHLWMPVPTVPGGSF
HFLWTFYVGRQGVQSGIPETISIRITAKRVCITGLHAVA VLGAASILLWTLAGAAYCT
ALLGRPLSIKORVNDNAAPFPFSKDDQGLMWLGIIPTNPDAEDLELKIGLITKNGI
PVKNVSDFEALQHNKAYVKLEIITGLNGEIRPDQRASYEGEHHELGLILFVKDDREDA
VASGS"
complement(8025. .9373)
/gene="yvkA"
complement(8025. .9359)
/gene="yvkA"
/notes="belongs to the major facilitator family"
/codon_start=1
/transl_table=11
/product="yvkA"
/protein_id="AAC67266.1"
/db_xref="GI:2618837"
/translation="MSSRKREKALIVLLICAILVUPINSTWIAVALSSISHTYNESIAS
ITWVTVYIIVMAVTOPIAGKGDYGNKTMYLWGVGLFLIASGLCALSPLLLIVF
RALQAVGGALTPNSITAIRHVVSEKRLPKVFGFGLGAGLGAALGPFIGSILIDFS
WHSIFWNTPFLAIALFTALTMTPEYKKNKSDAPLDIIGSLLAGSIVSILITKNEA

```

PMGVTYVSVLLVLPVFFRRKRTQHPIDFALFKSSTETNANLVLNLMMYAVL
LIMPLFTNFGNTSNMGLSVFISFMSASNNVGAQLHKKWGAKKIIFLSPAMAG
ANLLFLLSSSHSVLFLMLSLILGLASGVGLTSMOVSLATDPGMSGVASGIFSTF
RYFGSISSALIGLISGYHTLFMLIPAVSIIGVSVLGIKSDETARIEKNSA"
complement (9366. .9373)
/gene="yvkA"
complement (9390. .9972)
/gene="yvkB"
/note="9959"
complement (9390. .9959)
/gene="yvkB"
/note="belongs to the TetR/AcrR family of transcriptional
regulators"
/codon_start=1

Query Match 100.0%; Score 753; DB 1; Length 47739;
Best Local Similarity 100.0%; Pred. No. 5.3e-208;
Matches 753; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTGAATGAGATGACCGGAATGGTAAACGAAAGAGTCTGTGCAATTTTATGCTGAGGCA 60
Db 23747 GTGAATGAGATGACCGGAATGGTAAACGAAAGAGTCTGTGCAATTTTATGCTGAGGCA 23806
QY 61 TTAACAGAAACTGCAGAGAAATTTTGAACGGCGCAGGCATGTTTTGGTGGGATCAGC 120
Db 23807 TTAACAGAAACTGCAGAGAAATTTTGAACGGCGCAGGCATGTTTTGGTGGGATCAGC 23866
QY 121 CCATTTAAACAGCAGTTTTCAGAGGATATATTTACAGATTAATTTGATGGCGGAAAGCT 180
Db 23867 CCATTTAAACAGCAGTTTTCAGAGGATATATTTACAGATTAATTTGATGGCGGAAAGCT 23926
QY 181 CAATTTAAACAGCAGTTTTCAGTTTACTTTCAGGGGCATGAGCGGCTAACTTCTAGAACGC 240
Db 23927 CAATTTAAACAGCAGTTTTCAGTTTACTTTCAGGGGCATGAGCGGCTAACTTCTAGAACGC 23986
QY 241 CTTGGAACCTCGAGAGAAAGCTGAACGAAAGTAGAGAAAGAGGTATCAGAAACAGG 300
Db 23987 CTTGGAACCTCGAGAGAAAGCTGAACGAAAGTAGAGAAAGAGGTATCAGAAACAGG 24046
QY 301 AGATTTGAGAAAGAGCCCTTGTGGCTCATGCGGGGCATCGAAGGCGATTTCATACATTT 360
Db 24047 AGATTTGAGAAAGAGCCCTTGTGGCTCATGCGGGGCATCGAAGGCGATTTCATACATTT 24106
QY 361 TCTGATTTTATAGATAACAAAGCCTACAGCTGTGTGAGACAAGAGTTGAACATGCAATTT 420
Db 24107 TCTGATTTTATAGATAACAAAGCCTACAGCTGTGTGAGACAAGAGTTGAACATGCAATTT 24166
QY 421 TTTGAGCAGCCTCATTTTCAGCATGCTTGTGACATGCTCTGTGAAGCGATAATCGGG 480
Db 24167 TTTGAGCAGCCTCATTTTCAGCATGCTTGTGACATGCTCTGTGAAGCGATAATCGGG 24226
QY 481 CGTGGCGGGCGCTCAGTTTGATGATGAGAAAGTCAAGTCAAGTATGCTGAATTTGGCT 540
Db 24227 CGTGGCGGGCGCTCAGTTTGATGATGAGAAAGTCAAGTCAAGTATGCTGAATTTGGCT 24286
QY 541 GTGGAATATGCTAGTCAAGCTGCGCTTTTATCGGAGCTCGGATATTTAGAGGTG 600
Db 24287 GTGGAATATGCTAGTCAAGCTGCGCTTTTATCGGAGCTCGGATATTTAGAGGTG 24346
QY 601 GAAGAGACATCTCTTGTCTTATCATGCTGCGTGAAGCTGGGTGAGAAAGATCAGTAAACAT 660
Db 24347 GAAGAGACATCTCTTGTCTTATCATGCTGCGTGAAGCTGGGTGAGAAAGATCAGTAAACAT 24406
QY 661 GAATTTTCTATTTGATGCGGCGGAATCAAGGTGATCTCATTTGACAGGAATGGCGCAG 720
Db 24407 GAATTTTCTATTTGATGCGGCGGAATCAAGGTGATCTCATTTGACAGGAATGGCGCAG 24466
QY 721 ATGCTTTCTGAGAAACGGATCACATCTGAAGGA 753
Db 24467 ATGCTTTCTGAGAAACGGATCACATCTGAAGGA 24499

RESULT 2

BSUB0018/c
LOCUS
DEFINITION

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

BSUB0018 194692 bp DNA linear BCT 07-JUL-2003
Bacillus subtilis complete genome (section 18 of 21): from 3414339
to 3609030.
239121 AL009126
239121.2 GI:32468825
Bacillus subtilis subsp. subtilis str. 168
Bacillus subtilis subsp. subtilis str. 168
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
1 (bases 1 to 194692)
Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,
Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S.,
Boriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C.,
Bron, S., Brouillet, S., Bruschi, C.V., Caldwell, B., Capuano, V.,
Carter, N.M., Choi, S.K., Codani, J.J., Connerton, I.F., Cummings, N.J.,
Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D.,
Emmerson, P.T., Encian, K.D., Errington, J., Fabret, C., Ferrazi, E.,
Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A.,
Galleron, N., Ghim, S.Y., Glaser, P., Goffeau, A., Golightly, E.J.,
Grandi, G., Guisepi, G., Guy, B.J., Haga, K., Haiech, J., Harwood, C.R.,
Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F.,
Itaya, M., Jones, L., Joris, B., Karamata, D., Kaahara, Y.,
Klaer-Blanckhard, M., Klein, C., Kobayashi, Y., Koetter, P.,
Koningsstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Liu, H.,
Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H.,
Masuda, S., Maue, C., Medigue, C., Medina, N., Mellado, R.P.,
Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M.,
Ogawa, K., Ogiwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M.,
Portelle, D., Porwollik, S., Prescott, A.M., Presecan, E., Pujic, P.,
Purnelle, B., Rapoport, G., Ray, M., Reynolds, S., Rieger, M.,
Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadaie, Y., Sato, T.,
Scanlan, E., Schleicher, S., Schroeter, R., Scoffone, F., Sekiguchi, J.,
Sekowska, A., Seror, S.J., Serror, P., Shin, B.S., Soldo, B.,
Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K.,
Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpetra, P., Tognoni, A.,
Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A.,
Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzengraber, T.,
Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K.,
Yoshida, K., Yoshikawa, H.F., Zumstein, E., Yoshikawa, H. and
Danchin, A.

TITLE

JOURNAL
MEDLINE
PUBMED

REFERENCE
AUTHORS

TITLE
JOURNAL

Nature 390 (6657), 249-256 (1997)
93844033
9384377

2 (bases 1 to 194692)
Direct Submission

Submitted (27-JUN-2003) I. Moszer, A. Danchin, Institut Pasteur,
Genetique des Genomes Bacteriens, 28 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr.

adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
68 89 48

On Jul 7, 2003 this sequence version replaced gi:2635827.
This entry contains data from release R16.1 of the Subtilist
database. Further data on gene annotation and detailed information
about changes from previous releases can be found at
http://genolist.pasteur.fr/Subtilist/.

Location/Qualifiers
1. 194692
/organism="Bacillus subtilis subsp. subtilis str. 168"
/mol_type="genomic DNA"
/strain="168"
/db_xref="taxon:224308"
complement (79. .888)
/gene="fluC"
/locus_tag="BSU33290"
complement (79. .888)
/gene="fluC"
/locus_tag="BSU33290"
/codon_start=1
/transl_table=11

FEATURES
source

gene

CDS

Db 189088 CAATTTAAAGCGTTTCAGTTTTTACTTTGCGAGGCATGAGCGGCTAATCTTCTAGAGCG 189029
Qy 241 CTTTGGAACTCCGAGAGAAAGGCTGAACGAAAGTAGTAAAGAGGATATCAGCAAAACAGG 300
Db 189028 CTTTGGAACTCCGAGAGAAAGGCTGAACGAAAGTAGTAAAGAGGATATCAGCAAAACAGG 188969
Qy 301 AGATTTCAGAAAGAGCCCTTGTGGCTCATGCGGGGATCCGAAAGCGATATCATATTT 360
Db 188968 AGATTTCAGAAAGAGCCCTTGTGGCTCATGCGGGGATCCGAAAGCGATATCATATTT 188909
Qy 361 TCTGATTTATAGATAACAAAGCCTACACGCTGTGAGACAAGAGTTGAACATGCATTT 420
Db 188908 TCTGATTTATAGATAACAAAGCCTACACGCTGTGAGACAAGAGTTGAACATGCATTT 188849
Qy 421 TTTGAGCAGCCTCATTTTGCACATCTCTTGTGACATGCTCTGCTGAAGCGATAATCGGG 480
Db 188848 TTTGAGCAGCCTCATTTTGCACATCTCTTGTGACATGCTCTGCTGAAGCGATAATCGGG 188789
Qy 481 CGTGGCGGGCGCTCAGTTTGATGATGAAGAGAGTCACTGAGGATATGCTGAATTTGGCT 540
Db 188788 CGTGGCGGGCGCTCAGTTTGATGATGAAGAGAGTCACTGAGGATATGCTGAATTTGGCT 188729
Qy 541 GTGGAATATCTAGCTAGCTGAGCTCCGCTTTTATCGAGCTCCGGATATTTAGAGGTG 600
Db 188728 GTGGAATATCTAGCTAGCTGAGCTCCGCTTTTATCGAGCTCCGGATATTTAGAGGTG 188669
Qy 601 GAAGAGACACTCTTGTCTATCATCTGCTGGAAGCTGGGTGAGAGATCAGTAACCAT 660
Db 188668 GAAGAGACACTCTTGTCTATCATCTGCTGGAAGCTGGGTGAGAGATCAGTAACCAT 188609
Qy 661 GAATTTCTATTTGATGCGGCGCAATCAAGGATATCTATTGTACAGGAAATGGCGCAG 720
Db 188608 GAATTTCTATTTGATGCGGCGCAATCAAGGATATCTATTGTACAGGAAATGGCGCAG 188549
Qy 721 ATCTTTCTGAGAAACGATCATCTGAAGGA 753
Db 188548 ATGCTTTCTGAGAAACGATCATCTGAAGGA 188516

RESULT 3
LOCUS BX571859/c 349483 bp DNA linear BCT 26-SEP-2003
DEFINITION Photorhabdus luminescens subsp. laumondii TT01 complete genome;
segment 1/17.
ACCESSION BX571859 BX470251
VERSION BX571859.1 GI:36783455
KEYWORDS complete genome.
SOURCE Photorhabdus luminescens subsp. laumondii TT01
ORGANISM Photorhabdus luminescens subsp. laumondii TT01
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Photorhabdus.
REFERENCE 1
AUTHORS Duchaud,E., Rusniok,C., Frangeul,L., Buchrieser,C., Taourit,S.,
Bocs,S., Boursaux-Eude,C., Chandler,M., Dassa,E., Derose,R.,
Derzelle,S., Freysinet,G., Gaudriault,S., Givaudan,A., Glaser,P.,
Medigue,C., Lanois,A., Powell,K., Sigulier,P., Wingate,V.,
Zouine,M., Boenare,N., Danchin,A. and Kunst,F.
Complete genome sequence of the entomopathogenic bacterium
Photorhabdus luminescens
Nat. Biotechnol. 11 (1) (2003) In press
2
AUTHORS Duchaud,E., Frangeul,L., Rusniok,C. and Kunst,F.
Direct Submission
TITLE Submitted (23-APR-2003) L. Frangeul, Institut Pasteur, Genopole, 25
rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail:
lfrangeu@pasteur.fr, fkunst@pasteur.fr
LOCATION/Qualifiers
FEATURES
source 1. .349483
/organism="Photorhabdus luminescens subsp. laumondii TT01"
/mol_type="genomic DNA"
/strain="TT01"
/db_xref="taxon:243265"

gene 223. .1622
/gene="dnaA"
/locus_tag="plu0001"
223. .228
/gene="dnaA"
/locus_tag="plu0001"
234. .1622
/gene="dnaA"
/locus_tag="plu0001"
/codon_start=1
/transl_table=11
/product="Chromosomal replication initiator protein DnaA"
/protein_id="CAE12296.1"
/db_xref="GI:36783456"
/translation="MSLSLWQCLARLQDELPAFESMMIRPQLQABLNGNTLALYAPN
RFLVDVRSKYINNINGLNDYFCAGVPLLRFEVGNKPSQNDSPQRVVHTPTVAPAY
PONTVSPPSWMDTAVQPELSYRSNVNPKHTFDNFVEGKSNQLAARAQVADNPPGAY
NPLFLVGGTGLGKTHLLHAGNSIMARKANAKVVMHSEFVQDMVKALONNAIEEFK
RYRSVDALLIIDDIOFFANKERSQBEFFHTFALLEGNQOIIILTSRYPKEINGVEDR
LKSRFQWGTVAIEPPELETRVAIMKKADENEIQLPGEVAFIAKRLRNSVRELEGA
LNRVIANFTGRTAITDFVREALRDLLALQEKLVTDNIQKTVAEYIKTKVADLLSK
RRSRVARPRQWAMALAKELTNHSLPEIGDAFGGRDHTTVLHACRKIEQLREESHDIK
EDFSNLIRTLSS"
1627. .2727
/gene="dnaA"
/locus_tag="plu0002"
1627. .2727
/gene="dnaA"
/locus_tag="plu0002"
/codon_start=1
/transl_table=11
/product="DNA polymerase III, beta chain"
/protein_id="CAE12297.1"
/db_xref="GI:36783457"
/translation="MKFIITEREOLLKPLQVSSPLSGRPTLPILGNMLLLOVTEGSLLL
TGTDLFEMENMARVTLISLPHVGTATVPARKFDIWRGLPDGAEISVELDGDRLVRSR
RRRSLSTLPASDFPNLDWQSEVSELSQATLKRLIESTQFSMAHQDVRYLNGMLF
ETEGEELRTVADFTGHLAVCSMGIGQKLPSSHVSIVPRKGVIELMRLDGDGDTPLQLOI
GNNIRAHVQDFTFISKLVDGRPPDYRVLKPNPKDTLEANCMDLKAQFRAAILNSE
KFRGVRLYFSENLQIRITANNPEQEAEEIVDVSYQGAEMBEIGFNVSYVLVDNLTRCE
EVRLLLTDAISSVQIIEDCANHNAAYVVMPEML"
2749. .3840
/gene="recF"
/locus_tag="plu0003"
2749. .3840
/gene="recF"
/locus_tag="plu0003"
/codon_start=1
/transl_table=11
/product="DNA replication and repair protein"
/protein_id="CAE12298.1"
/db_xref="GI:36783458"
/translation="MTLRLIRDFENIAAADLPLATGFNFLVNSGSGKTSVLEATY
TIGHGSRFSIOAGRVILHGCDFFVLHGRLOQENERSISIGLSKRNKSGSKVIDGS
DGKTIABLAMLPQMLITPEGFTLLNGPKYRFAIDWGCFFHNEPFPFSAWNLKRL
KORNAALRVTRYQSIRPMDQELIPLANQINWRGYVNTIQTIDNTCKQFLPEFTL
SFSFQCGWCKDESDYALRQEFERDRTLTATSGPHKADLRIRAEGETPVEDMLSRGOL
KLIMCALRAQGEYFTFQSGOQCLYLLDDFASELDTSRRQLLAARLKSTQAQVFSAI
NPQDITDMLDGNKMFVRVENGKIEVQPD"
3860. .6326
/gene="gyrB"
/locus_tag="plu0004"
3860. .6274
/gene="gyrB"
/locus_tag="plu0004"
/codon_start=1
/transl_table=11
/product="DNA gyrase subunit B"
/protein_id="CAE12299.1"
/db_xref="GI:36783459"
/translation="MSNTYDSSSIKVLKGLDAVRKPPGMWIGTDDGTGLHWMFEVW
DNAIDBALAGHCSEVITVTHADNSVQDDGRGIPGTIHEEGVSAAEVINTVLHAGG
KFDNDSKYVSGGLHGVSVVNALSEKLVITRRDGKVHEQTTHLGVFPQSPKLVVGET

```

EOTGTRVRFPMSMDTFRDNTFQHDILAKRLRELSFLNSGVSLRLDKRTNIBDHPHY
EGGKAFVFLNKKTPIHENVFYFSTEKGIGVEISLMOVNDGFOENIYCFNTNIPOR
DGGTHLVFRTATNRTLSYMDKEGYNKKSKVSATGDDAREGLAVISVVPDPKFFSS
QTKDLVSSEVKTAVETLMNEKLVYELLENPNDAKTUVSKIIDAAARAARAKAREMT
RRKGALDLAGLPQKQDCQDRDPALBELYLVEGDSAGGSAKQGRNKNQAIPLPKGKI
LNVKARFDMKLSQSEVATLITAGCGIGRDEYNPKLRYSIIIMTDADVDSHIRT
LLITFFYRQPEIIERGHVFIAQPPLYKVKKGQEQVIKQDEAMDEYQLSIDGASL
YTNHVPALRGEPLEKLVTFNGVQKIIKRMERLYPLSLNLSLIYHPNLTEETLSDKA
KVHEMLEMLVLRDLNDEONGSSYSYBLHENRERQVPEVLCVPTHGVDYDILLDFPV
HGGEYHRITLLGDKRLNLEEDAYIERGERRQVPSFEQALNWLTKESRGLSIOQRYK
GLGEMPEQJLWETMNPETRMVRVTYKQAIATDQLFTLLMGDAVEPRAFIEENALK
AANIDI".6326
/terminator
/genes="gyrB"
/locus_tag="plu0004"
6415..6926
/locus_tag="plu0005"
6415..6894
/locus_tag="plu0005"
/function="Unknown"
/notes="unnamed protein product; Similar to Hcp protein"
/codon_start=1
/transl_table=11
/protein_id="CAE12300.1"
/db_xref="GI:36783460"
/translations="MADMIYMTIKGKQGLISAGCSSVDSIGNKYQANHFDQILVYSL
SHAITQNVHDQPIIIQKPIKSSPLLGVAISENSLECSIDPYETNSNGAQERYN
IKITGATITRISIVLHPSLHNHLELPQESISLKYITNWNHLIAGTGGYSIMDDRYY"
6908..6926
/locus_tag="plu0005"
/complement(8013..8778)
/locus_tag="plu0006"
/complement(8013..8768)
/locus_tag="plu0006"
/function="Unknown"
/notes="unnamed protein product; Similar to unknown
protein"
/codon_start=1
/transl_table=11
/protein_id="CAE12301.1"
/db_xref="GI:36783461"
/translations="MKQDNLDDLVKRELRELAKTLSFRVLTHSKIRMDFNKINLFVE
DILOQVIRHCLSNLGAITEIIOFEDHLEKEOAFCLTANRVKQYAIIEREKESYTNLI
LKQIFVGGGTQIAGTVYCKASIGLACASFGAPLMAHGYNVNVENGYYLLYRENING
GVREGRYVANKLGLSDKADITYATVDLVLSGIGIFRKYKLPREKSWSLFRNINSDF
IRWREMGVSVLITTEGIVDVSVLSIYQLEKEK"
/locus_tag="plu0006"
/complement(8773..8778)
/locus_tag="plu0006"
/complement(8935..10056)
/genes="asd"
/locus_tag="plu0007"
/complement(8935..10044)
/genes="asd"
/locus_tag="plu0007"
/codon_start=1
/transl_table=11
/product="aspartate-semialdehyde dehydrogenase (ASA
dehydrogenase) (ASADH)"
/protein_id="CAE12302.1"
/db_xref="GI:36783462"
/translations="MKNVGFIGWRGMVGSVMQRMIEEDFDVICPVFFTTSHQGP
APDFTGOQQLQNAFDIHALGALDIIISCOGGDYTNEIYPLKRAATGQGYWIDAASAL
RMNDSIILIDPVNHTHIOQLNGIKITFVGNGCTVSLMLSLGGIPANDLVWASVA
TYQASGARHRELIVQMSLHTQVAKELQDPASAILDIERKVDTPFRSGVNSTEK
FSVPLAGSLIPWIDKQDNGOSREEMWKGQAEINILNTGNIIITVDGLCVRIGALRCH
SQAPTLKXLDISIPETIEQLAAHNDWVRVIPNDRELMSRELTPAAVTTGLDTPVGR
KLNMGPEVLSAFTVQQLLWGAAPLRLMLRILI"
/complement(10051..10056)
/genes="asd"
/locus_tag="plu0007"
/complement(10251..10472)
/genes="ogrK"
/locus_tag="plu0008"

```

```

complement(10251..10472)
/genes="ogrK"
/locus_tag="plu0008"
/function="Unknown, probable positive regulator of phage
P2 late gene transcription"
/codon_start=1
/transl_table=11
/product="Ogr protein"
/protein_id="CAE12303.1"
/db_xref="GI:36783463"
/translations="MIKPLCCGAHAHARSFHSCHTKERYNQCQININGCATFVSHET
FYRFSKPGEVISVXPKPKKTKTQLNVV"
/complement(10549..11634)
/locus_tag="plu0009"
/complement(10549..11634)
/locus_tag="plu0009"
/locus_tag="plu0009"
/function="Unknown, probable tail synthesis"
/notes="unnamed protein product; Similar to protein D (gpd)
of Enterobacteria phage P2"
/codon_start=1
/transl_table=11
/protein_id="CAE12304.1"
/db_xref="GI:36783464"
Query Match 8.2%; Score 61.6; DB 1; Length 349483;
Best Local Similarity 46.4%; Pred. No. 3.9e-06;
Matches 310; Conservative 0; Mismatches 349; Indels 9; Gaps 3;
QY 46 TTTATTGCTGAGGCATTAACAGAAAACCTGCAGAGAAATATTTGAACGGCGCAGGCATGTT 105
DB 315652 TTTACTGTCCAAAGGTGAAACCTCTCGTTGTGACCAATATTATCAAAAAGGTGATCATCGCG 315593
QY 106 TTGGTGGGATCAGCCCATTTAACAGCAGGTTTTACAGAGATTTATTTACAGATTAATT 165
DB 315592 CTAATAGGATAAGCCCTTTTAACTCGCGTTTTTCAAAAAGACTATGTAGTGGACCTTATT 315533
QY 166 GGATGGGCGAAAGCTCAATT---TAAAGCGTTTTTCAGTTTTTACTTCAGGGCATGAGCGC 222
DB 315532 CAGTGGTCAAGTCATTTATTTCCGACAGTCGACATATTATTACCTTGTGAACGTGAAGCT 315473
QY 223 GCTAATCTTCTAGAACGGTTGGAACCTCCGAGAGGAAAGCTGAAACGAAAGTAAGGAAA 282
DB 315472 TCACGCGCTTTTAGTCGTAGTGAATGATAATGTTTAAAGCTATCAAAAACACATCGC 315413
QY 283 GAGGTATCAGAAACAGGAGATTTCAGAAAGAGCCCTGTGCGC----TCATGGCGGGG 337
DB 315412 GAAATGTAGCGTCATTTACGTAACTTGATTTATGTTATTTCCACAGCAACATTGAAAAGT 315353
QY 338 ATCCGAAGCGATTTCATAC-ATTTTCTGATTTTATAGATAACAAAGCTACCAGCTGTTG 396
DB 315352 AAGCAATCAGAGTCATCCAAATTTAGTGACTTTTCACTAAACCATGACTACCAATCTCT 315293
QY 397 AGACAAAGAGTTGAACATGCATTTTGTGAGCAGCCTCATTTTCGACATGCTTGTGTTGGAC 456
DB 315292 AAAACACAAGTTGAAAACGCGTTTAATGAATCAGAAATCTTTTAAAAAAGCTGTCTTGAT 315233
QY 457 ATGCTCTCGTAGCGATAATCGGGCGTCGCGGGCGCTCAGTTTGTATGATGAGGAAGATC 516
DB 315232 ATGCTCTTCAAGCCATAAAGGGCGCACTAAAGAGTACTGGGCAATACCTTTTGTGCAAA 315173
QY 517 AGTGAGGATATGCTGAATTTGGCTGTGGAAATATGTTCATAGCTGAGCTGCGGCTTTTTTATC 576
DB 315172 GACCTACAATTAGTATATAAAGCGTTGCCATATATTTCGTGAAATTCCTTTTACCTC 315113
QY 577 GGAGCTCCGGATATTTTAGAGGTGGAAGAGACACTCCTCTGCTTATCATTCGTCGTTGAAG 636
DB 315112 AATACCCCTCGATTACTTGGGGTAAAGTATTTCTACGTTACTTTATCACCCCTTGGTCA 315053
QY 637 CTGGGTGAGAAGATCAGTAACCATGAATTTTCTATTGTTATGCGCGCAATCAAGGGTAT 696
DB 315052 ATCGGAAAAGGGTTATTTAACGGTAGTTATTCCTATACAGTAGCAGATAAACAAGTTAC 314993
QY 697 CTCATTGT 704

```



```
Db 314992 GGAATCGT 314985
|||
RESULT 4
AX770907 349980 bp DNA linear PAT 02-JUL-2003
LOCUS
DEFINITION Sequence 38 from Patent WO02094867.
ACCESSION AX770907
VERSION AX770907.1 GI:32438071
KEYWORDS
SOURCE
ORGANISM
Photorhabdus luminescens
Photorhabdus luminescens
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Photorhabdus.
1
REFERENCE
Duchaud, E., Taurit, S., Glaser, P., Frangeul, L., Kunst, F.,
Danchin, A. and Buchrieser, C.
TITLE Sequence of the Photorhabdus luminescens strain TT01 genome and
uses
JOURNAL Patent: WO 02094867-A 38 28-NOV-2002;
INSTITUT PASTEUR (FR); CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE (CNRS) (FR)
FEATURES
source
1. 349980
/organism="Photorhabdus luminescens"
/mol_type="genomic DNA"
/db_xref="taxon:29488"
/notes="sequence length to big . Splitted into 2
sequences. -seq 00038: 1 too 349980-seq 10787: 300001 to
618776"
ORIGIN
Query Match 8.2%; Score 61.6; DB 6; Length 349980;
Best Local Similarity 46.4%; Pred. No. 3.9e-06;
Matches 310; Conservative 0; Mismatches 349; Indels 9; Gaps 3;
Qy 46 TTTATTGCTCAGGCATTAAACAGAAACTGCAGAGAAATATTGACGGCGCAGGCATGTT 105
Db 282443 TTTACTGTCGAAGTGAACTCTCGTTGTGACCAATTTTCAAAAAGTGATCAGCG 282502
Qy 106 TTGGTGGGATCAGCCATTAAACAGCAGGTTTTCAGAGATATATTACAGATTAATT 165
Db 282503 CTAATAGGAGTAAGCCCTTTAACTCGCGTTTTTCAAAAAGCATGTAGTGGACCTTATT 282562
Qy 166 GGATGGCGGAAAGCTCAATT---TAAAGCGTTTCAGTTTACTTTCAGGCGCATGAGCGG 222
Db 282563 CAGTGGTCAAGTCNATTTTCCGCAAGTGCACATATATTACCTTGTGAACGTGAAGCT 282622
Qy 223 GCTAATCTCTAGAAGCGCTTGGAACTCCGAGAGGAAAGGCTGAACGAAAGTAAGGAA 282
Db 282623 TCAGCGCTTTTAGTCGTAGTGAATGATAATGTTAAAGCTATCAAAAACACATCGC 282682
Qy 283 GAGGTATCAGAAACAGGAGATTGCGAGAAAGCCCTTGTGGC-----TCATGGCGGGG 337
Db 282683 GAAATTAGACGTCATTACGTAACCTTGATTATGTTATTTCCACAGCAACATTGAAAGT 282742
Qy 338 ATCGAGAGCGGNTTCATAC-ATTTCTGATTTTATAGATACAAAGCCTACCGCTGTG 396
Db 282743 AAGCAATCAGATCATCCAAATTTAGTGACTTTTCACTAAACCATGACTACCAATCTCTT 282802
Qy 397 AGACAAGAAGTTGAACATGCATTTTTTTGAGCAGCGCTCATTTTCGACATGCTTGTGTCAC 456
Db 282803 AAAACACAAGTTGAAACGCGTTTAATGAATCAGAACTTTTAAAAAAGCTGCTTGAT 282862
Qy 457 ATGTCTCGTGAAGGATAATCGGCGGTGCGCGGGCGTCAAGTTTGATGATGAAGAAGTC 516
Db 282863 ATGTCTCTTCAAGCATAAAGGGCGACTAAAAGGTACTGGCAATACTTTGGTCAAAATT 282922
Qy 517 AGTGAGATATGCTGAATTTGGCTGTGGAATATGTCATAGTCAAGCTGCGCTTTTATC 576
Db 282923 GACCTACAATTAGTATATAAGCGGTGCCATATATTTTCGTGAAATTCCTTTTACCTC 282982
```

```
Qy 577 GGAGCTCCGGATATTTTAGAGGTGGAAGACACACTCTTCTGCTTATCATCGTCCGTGAAG 636
Db 282983 AATACCCCTCGATTACTTGGGTAAAGTATTTCTACGTACTTATCATCCGCCCTTGTCA 283042
Qy 637 CTGGGTGAGAAGATCAGTACCATGAATTTTCTATTGTTATCGGCCGGAATCAAGGTAT 696
Db 283043 ATCGAAAAGGGTTATTTAACGGTAGTTATCTTATACAGTACGAGATAACAAGTTAC 283102
Qy 697 CTCATTGT 704
Db 283103 GGAATCGT 283110
RESULT 5
I66494/c
LOCUS
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION I66494
VERSION I66494.1 GI:2724471
KEYWORDS
SOURCE
ORGANISM
Unknown.
Unclassified.
REFERENCE
1 (bases 1 to 7218)
AUTHORS Dörner, F., Scheiflinger, F. and Falkner, F. Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
Location/Qualifiers
1. 7218
source
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN
Query Match 6.5%; Score 48.6; DB 6; Length 7218;
Best Local Similarity 4.3%; Pred. No. 0.012;
Matches 15; Conservative 195; Mismatches 139; Indels 0; Gaps 0;
Qy 1 GTCAATGAGATGACCGGAATGGTAAACGAAAGGTCTGTGCATTTTATTTGCTGAGGCA 60
Db 1460 GTTAAAGAGATAGAAGAAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1401
Qy 61 TTAACAGAAACTGCAGAGAAATATTTGAACGGCGCAGGCATGTTTGTGGGGATCAGC 120
Db 1400 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1341
Qy 121 CCATTAAACAGCAGGTTTTCAGAGGATATATTACAGATTAATTGATCGCGGAAAGCT 180
Db 1340 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1281
Qy 181 CAATTTAAAGCGCTTTCAGTTTTCAGGGCATGAGCGGCTAATCTTCTAGAAAGC 240
Db 1280 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1221
Qy 241 CTTTGAACCTCCGAGAGAAAGGCTGAACGAAAGTAAGGAAAGAGGTATCAGGAAACAG 300
Db 1220 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1161
Qy 301 AGATTTCAGAAAGACCCCTTGTGGCTCATGGCGGATCCGAGGCGA 349
Db 1160 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1112
RESULT 6
AC040918/c
LOCUS
DEFINITION Homo sapiens chromosome 15, clone RP11-684B21, complete sequence.
ACCESSION AC040918
VERSION AC040918.7 GI:18693519
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

1 (bases 1 to 175919)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-684B21
Unpublished
2 (bases 1 to 175919)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Buckett,G.,
Campiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Olivier,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Subramanian,N., Subramanian,A., Talamas,J.,
Tessfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

3 (bases 1 to 175919)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collimore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tessfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (01-SEP-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

4 (bases 1 to 175919)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collimore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C.,
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,

Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tessfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (16-FEB-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 16, 2002 this sequence version replaced gi:15412492.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8454
Center clone name: 684_B_21

FEATURES
source

1. .175919
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone_lib="RP11-684B21"
/clone="RP11-684B21"
complement(1199..1503)
/rpt_family="MER7A"
complement(1776..1905)
/rpt_family="LIME"
1906..2388
/rpt_family="MLT1J"
complement(2389..2451)
/rpt_family="LIME"
complement(2552..2624)
/rpt_family="MIR"
4152..4300
/rpt_family="MIR3"
4651..4748
/rpt_family="LIPAS"
complement(5313..5377)
/rpt_family="L2"
complement(5916..6051)
/rpt_family="MIR"
6227..6698
/rpt_family="MER70A"
6783..6810
/rpt_family="AT_rich"
6920..6987
/rpt_family="(TCTA)n"
complement(8931..9079)
/rpt_family="MIR"
9339..9586
/rpt_family="L2"
9625..9698
/rpt_family="L2"
9788..10304
/rpt_family="LIMB2"
10310..10674
/rpt_family="LIMB4"
11089..11296
/rpt_family="AluY"
complement(11356..11558)
/rpt_family="MIR"
12004..12024
/rpt_family="AT_rich"
13583..13621
/rpt_family="(CATA)n"
complement(14795..14940)

PUBMED 10381325
REMARK Submitted to Molecular Phylogenetics and Evolution
REFERENCE 2 (bases 1 to 1041)
AUTHORS Omland, K.E., Lanyon, S.M. and Fritz, S.J.
TITLE Direct Submission
JOURNAL Submitted (19-OCT-1998) Bell Museum of Natural History, University of Minnesota, 1987 Upper Buford Circle, Saint Paul, Minnesota 55108, USA

FEATURES
source Location/Qualifiers
1..1041
/organism="Amblycercus holosericeus"
/organella="mitochondrion"
/mol_type="genomic DNA"
/db_xref="taxon:84791"
1..1041
/notes="NADH-2"
/codon_start=1
/transl_table=2
/product="NADH dehydrogenase subunit 2"
/protein_id="AAF03308.1"
/db_xref="GI:6090722"
/translation="MPRASLIETTSLLGLTTITSSNHNWMAWTGLEINTLAILPLI
SKSHPRAPAEAAKTYLTOASALVLFSSMTNMTGQWDITQLTHPTSCLLTSAI
AMKLGVPFHFPEVLQGSPLTGLLSTIMKLPITLLYMTSPSLNPTLLTILAIL
STALCGWMLNTOIRKILAFSSISHLWMAIIIVNPKLTILFYLXAMMTLIELT
LNTIKVLKLTMTAKTSPSLNMLLTLISLAGLPPLTGPLPKLIIQELTKQMA
PAATLIALLSLLSFYRLAYCATITLPHITNHNKQWRTNKPNTITAILITVSA
LLPISPMILTFI"

ORIGIN
Query Match 5.7%; Score 42.8; DB 5; Length 1041;
Best Local Similarity 54.4%; Pred. No. 0.44; Indels 0; Gaps 0;
Matches 86; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
QY 585 GGAATATTTAGAGTGAAGAGACACTCTTCCTTATCATCGTCGCGGAAGCTGGTGA 644
DB 603 GGTGAGTTTGGTGTGATGATGATAATGATGCTATCATCTAGTGGGAGATGATGA 544
QY 645 GAGATCAGTAACGAAATTTCTATTGTATGCGGCCGAATCAAGGGTATCTCATGTT 704
DB 543 AAAGGCTAGGATTTTCGGATTTGTGTTGTTGAGGCCCATTCACACCTCCGAGGGCTGT 484
QY 705 ACAGGAATCGCGCAGATGCTTTCTGAGAAACGGATCA 742
DB 483 AGAGAGATCGCTAAGTGGTTAATAGTGTGGGGTTCA 446

RESULT 9
E08319
LOCUS E08319 cDNA encoding luciferase from firefly.
DEFINITION E08319
ACCESSION E08319.1 GI:2176437
VERSION E08319.1
KEYWORDS JP 1994030982-A/1.
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 1659)
AUTHORS Yoshino, S., Shiraishi, S., Inoue, S. and Saigo, K.
TITLE FIREFLY ENZYME LUCIFERASE GENE
JOURNAL Patent: JP 1994030982-A 1 01-NOV-1994;
CHISSO CORP
COMMENT OS Photuris (firefly)
PN JP 1994030982-A/1
PD 01-NOV-1994
PF 21-APR-1993 JP 19931119050
PI YOSHINO SHUHEI, SHIRAISHI SHINJI, INOUE SATOSHI, SAIGO KAORU
PC C12N15/53, C12N1/21, C12N9/02, C12N15/70, (C12N1/21, C12R1.19), PC (C12N9/02,
PC C12R1.19);
CC strandedness: Single;
CC topology: Linear;
CC hypothetical: No;

anti-sense: No; Location/Qualifiers
Key source 1..1659
/organism="Photuris"
/clones="ppFL7"
1..1659
/product="Luciferase from firefly".

FEATURES
source Location/Qualifiers
1..1659
/organism="unidentified"
/mol_type="genomic RNA"
/db_xref="taxon:32644"

ORIGIN
Query Match 5.6%; Score 41.8; DB 6; Length 1659;
Best Local Similarity 29.3%; Pred. No. 0.94;
Matches 90; Conservative 33; Mismatches 184; Indels 0; Gaps 0;
QY 307 GCAGAAAGAGCCCTTGCTCATGCGGGGATCCGAGCGGATTCATACATTTCTGAT 366
DB 1258 GNGAYATGGNTAYTTYGAYGARGAYGNCAYGTTATATHTGTNGAYNGNNTNAARNN 1317
QY 367 TTTATAGATAACAAGCCCTACCGAGCTGTTTGAGACAAAGAGTTGAACATGCAATTTTGGAG 426
DB 1318 NTNATHAARTAYAARGNTAYCARGTNCNCNGCNGARNTNGARGCNCNNNTNNTNCAR 1377
QY 427 CAGCCTCATTTTCGACATGCTGTTTGGACATGCTCTCGTGAACGATATATCGGCGGTGCG 486
DB 1378 CAYCNCNTTATGARGAYGNGGNGTNGCNGGNTCCNGAYGARGTNGCNGNGAYNTN 1437
QY 487 CGGGGCGTCAGTTTGATGATGGAAGATCAGTGAGGATATGCTGAATTTGGCTGTGGAA 546
DB 1438 CNGGNGCNGTNGTNGTNTNNAARGGNAARNNNATACNGARAARGARATHCARGAY 1497
QY 547 TATGTCATAGCTGAGCTGCCGTTTATTCGGAGCTCCGGATATTTTAGAGGTGGAAGAG 606
DB 1498 TAYGTNGCNGCARGTNCACNNNNNNNAARAENTNNGGNGNGTNGARTTYGTNAAR 1557
QY 607 ACACTCC 613
DB 1558 GARGTNC 1564

RESULT 10
AF526473/c
LOCUS AF526473 Motacilla flaviiventris specimen-voucher FMNH352834 NADH
DEFINITION dehydrogenase subunit 2 (ND2) gene, partial cds; mitochondrial gene for mitochondrial product.
ACCESSION AF526473
VERSION AF526473.1 GI:24899324
KEYWORDS
SOURCE
ORGANISM mitochondrion Motacilla flaviiventris (Madagascar wagtail)
Motacilla flaviiventris
Archosauaria; Aves; Neognathae; Passeriformes; Motacillidae;
Motacilla.
REFERENCE 1 (bases 1 to 1006)
AUTHORS Voelker, G.
TITLE Systematics and Historical Biogeography of Wagtails: Dispersal Versus Vicariance Revisited
JOURNAL Condor 104 (4), 725-739 (2002)
REFERENCE 2 (bases 1 to 1006)
AUTHORS Voelker, G.
TITLE Direct Submission
JOURNAL Submitted (02-JUL-2002) Barrick Museum of Natural History, University of Nevada Las Vegas, 4505 Maryland Parkway, Las Vegas, NV 89154, USA

FEATURES
source Location/Qualifiers
1..1006
/organism="Motacilla flaviiventris"
/organella="mitochondrion"

```
/mol_type="genomic DNA"
/specimen_voucher="FMNH352834"
/db_xref="taxon:211421"
<1..>1006
/gene="ND2"
CDS
<1..>1006
/gene="ND2"
/codon_start=1
/transl_table=2
/product="NADH dehydrogenase subunit 2"
/protein_id="AAN64961.1"
/db_xref="GI:24899325"
/translation="AKLIPTISLLGTTITISSNHWIMAWAGLEINTLAILPLISKSH
HPRAEATKYFLVQAAASALVLFSSMTNAWCTGQWDITQLTHPTSLCVLTSAIAMKL
GLVPHFWFPEVLQSGPLTGLLSTAMKLPITLLYMTSPSLNPTLLTMTAILSTAL
GWMGLNQTOQIRKILAFSSISHLGWMALILTYPNKLTLNLYFYALYMTTFTMTSM
KVLKLSLTMTAWTKVPSLNAMLLLTLLSLAGLPPLTGFLPKWLIIOBLTKODMAPAT
IISLLSLSLFFYLRLAYCTTITLPHSTNHWIMQWIIHKPTPTLAILITVSVTLPA
S"
```

ORIGIN

```
Query Match 5.5%; Score 41.2; DB 5; Length 1006;
Best Local Similarity 53.8%; Pred. No. 1.3;
Matches 85; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 585 GGATATTTTACAGTGAAGAGACACTCTCTGCTTATCATCGTCCGCGGAGCGGTGA 644
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 645 GAAGATCAGTAACCATGAATTTCTATTGTCGCGCGGAGGATCTCATTTGT 704
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 705 ACAGGAATGCGGAGATGCTTCTGAGAAACGATCA 742
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 471 TGAAGATGCTATGCTGTTAATAGTAGTGGTTCA 434
```

RESULT 11

```
AF526481/c
LOCUS
DEFINITION
Motacilla flaviiventris specimen-voucher FMNH352832 NADH
dehydrogenase subunit 2 (ND2) gene, partial cds; mitochondrial gene
for mitochondrial product.
```

ACCESSION

AF526481

VERSION

AF526481.1 GI:24899340

KEYWORDS

mitochondrion Motacilla flaviiventris (Madagascar wagtail)

SOURCE

Motacilla flaviiventris

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

Voelker, G.

AUTHORS

Systematics and Historical Biogeography of Wagtails: Dispersal

TITLE

Versus Vicariance Revisited

JOURNAL

Condor 104 (4), 725-739 (2002)

REFERENCE

2 (bases 1 to 1006)

AUTHORS

Voelker, G.

TITLE

Direct Submission

JOURNAL

Submitted (02-JUL-2002) Barrick Museum of Natural History,
University of Nevada Las Vegas, 4505 Maryland Parkway, Las Vegas,
NV 89154, USA

FEATURES

Location/Qualifiers

1..1006

/organism="Motacilla flaviiventris"

/organelle="mitochondrion"

/mol_type="genomic DNA"

/specimen_voucher="FMNH352832"

/db_xref="taxon:211421"

<1..>1006

/gene="ND2"

<1..>1006

```
/gene="ND2"
/codon_start=1
/transl_table=2
/product="NADH dehydrogenase subunit 2"
/protein_id="AAN64969.1"
/db_xref="GI:24899341"
/translation="AKLIPTISLLGTTITISSNHWIMAWAGLEINTLAILPLISKSH
HPRAEATKYFLVQAAASALVLFSSMTNAWCTGQWDITQLTHPTSLCVLTSAIAMKL
GLVPHFWFPEVLQSGPLTGLLSTAMKLPITLLYMTSPSLNPTLLTMTAILSTAL
GWMGLNQTOQIRKILAFSSISHLGWMALILTYPNKLTLNLYFYALYMTTFTMTSM
KVLKLSLTMTAWTKVPSLNAMLLLTLLSLAGLPPLTGFLPKWLIIOBLTKODMAPAT
IISLLSLSLFFYLRLAYCTTITLPHSTNHWIMQWIIHKPTPTLAILITVSVTLPA
S"
```

ORIGIN

```
Query Match 5.5%; Score 41.2; DB 5; Length 1006;
Best Local Similarity 53.8%; Pred. No. 1.3;
Matches 85; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 585 GGATATTTTACAGTGAAGAGACACTCTCTGCTTATCATCGTCCGCGGAGCGGTGA 644
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 645 GAAGATCAGTAACCATGAATTTCTATTGTCGCGCGGAGGATCTCATTTGT 704
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 705 ACAGGAATGCGGAGATGCTTCTGAGAAACGATCA 742
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 471 TGAAGATGCTATGCTGTTAATAGTAGTGGTTCA 434
```

RESULT 12

```
I40164
LOCUS
DEFINITION
Sequence 1 from patent US 5618722.
ACCESSION
I40164
VERSION
I40164.1 GI:2083169
KEYWORDS
Unknown.
SOURCE
Unknown.
ORGANISM
Unclassified.
```

REFERENCE 1 (bases 1 to 1659)

AUTHORS Zenko, S., Shiraishi, S., Inouye, S. and Saigo, K.

TITLE Photuris firefly luciferase gene

JOURNAL Patent: US 5618722-A 1 08-APR-1997;

FEATURES Location/Qualifiers

source

1..1659

/organism="unknown"

/mol_type="unassigned DNA"

ORIGIN

Query Match 5.5%; Score 41.2; DB 6; Length 1659;

Best Local Similarity 30.0%; Pred. No. 1.4;

Matches 93; Conservative 39; Mismatches 178; Indels 0; Gaps 0;

QY 304 TTTCAGAAAGAGCCCTTGCTGCTCATGCGCGGATCCGAGGCGATTCATCATTTCT 363

Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1255 TCNGNGAYATHGNTAYTYGAYGARGAYGNCAYGTNTAYATHGNTGAYMGNTNAR 1314

Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 364 GATTATTAGATAACAAAGCCTACAGCTGTTGAGCAAGAGTGTGAACATGCAATTTT 423

Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1315 TCNTNATHAARTAYAAAGGNTAYCARGTNCNCNCNGAYGTNGAGCNGNTNYNTN 1374

Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 424 GAGCAGCCTCATTTTCGACATGCTTGTGTCATGTCGTCGAGGATGCTGTAATCGGCGT 483

Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1375 CACAYCCNTTYATHGARGAYCNGNGTNCNGAYGARGTNGCNGNGAY 1434

Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 484 GCAGCGGCGCTCAGTTTGTATGATGGAAGAGTCACTGAGGATGCTGTAATTTGGCTGTG 543

Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1435 YTNCCNGCNGCNGTNGTNGTNYNNAARGAGGNAATCNATHACNGARARAGARATHCAR 1494

Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 544 GAATATGTATAGCTGAGCTGCCGCTTTTATCGGAGCTCCGGATATTTTAGAGGTGAA 603

Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```
Db 1495 GAYTAYTGNGCNGCARGTACACGCTGCTCNAARAARYTNMGNGGNGGTNGARTTYGTN 1554
Qy 604 GAGACACTCC 613
Db 1555 AARGARGTNC 1564

RESULT 13
LOCUS AY294513/c 1659 bp DNA linear PAT 13-MAY-1997
DEFINITION Sequence 2 from patent US 5618722.
ACCESSION I40165
VERSION I40165.1 GI:2083170
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1659)
AUTHORS Zenko,S., Shiraishi,S., Inouye,S. and Saigo,K.
TITLE Photuris firefly luciferase gene
JOURNAL Patent: US 5618722-A 2 08-APR-1997;
FEATURES
source Location/Qualifiers
1..1659
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 5.5%; Score 41.2; DB 6; Length 1659;
Best Local Similarity 30.0%; Pred. No. 1.4;
Matches 93; Conservative 39; Mismatches 178; Indels 0; Gaps 0;

Qy 304 TTTCGAGAAAGAGCCCTTGGCTCATGCGGGGATCCGAGGGGATTCAATTTCT 363
Db 1255 TCNGGNGAYATHGGNTAYTTTGAYGARGAYGNCAYGNTAYATHGNGAYMGNYNAAR 1314
Qy 364 GATTTATAGATACAAAGCCTACCAGCTGTGAGACAAGATGGAACATGCATTTT 423
Db 1315 TCNTYNATHAATAAAGNTAYCARGTCCNCNCNGARYTNGARGCNYNTYN 1374
Qy 424 GAGCAGCCTCATTTTCACATGCTGTTTGGACATGCTCTGCGAAGCGATAATCGGCGCT 483
Db 1375 CARCAYCNTTAYATHGARGAYGCGNGTNGCNGGNTCCNGAYGARGTNGCNGNGAY 1434
Qy 484 GCGGGGGCGTCAGTTGATGATGGAAGAAGTCACTGAGGATATCGTAATTTGGCTGTG 543
Db 1435 YTNCCNGGCGTNGTNGTNGTNYNAARGAGGNAARTCNATHACNGARAARGARATHCAR 1494
Qy 544 GAATATGTCATAGCTAGCTGAGCTGCCGCTTTTATCGGAGCTCCGGATATTTTAGAGGTGAA 603
Db 1495 GAYTAYTGNGCNGCARGTACACGCTCNAARAARYTNMGNGGNGGTNGARTTYGTN 1554

Qy 604 GAGACACTCC 613
Db 1555 AARGARGTNC 1564

RESULT 14
LOCUS AY294513/c 1056 bp DNA linear VRT 28-FEB-2004
DEFINITION Aerodramus terraereginae specimen-voucher DHC20 NADH dehydrogenase
subunit 2 (ND2) gene, partial cds; mitochondrial.
ACCESSION AY294513
VERSION AY294513.1 GI:37812896
KEYWORDS
SOURCE Mitochondrion Aerodramus terraereginae
ORGANISM Aerodramus terraereginae
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Apodiformes; Apodidae; Aerodramus.
1 (bases 1 to 1056)
PRICE,J.J., Johnson,K.P. and Clayton,D.H.
The evolution of echolocation in swiftlets
J. Avian Biol. 35 (2), 135-143 (2004)

REFERENCE
AUTHORS Price,J.J., Johnson,K.P. and Clayton,D.H.
TITLE The evolution of echolocation in swiftlets
JOURNAL J. Avian Biol. 35 (2), 135-143 (2004)
FEATURES
source Location/Qualifiers
1..1056
/organism="Aerodramus vanikorensis"
/mol_type="mitochondrial"

ORIGIN
Query Match 5.4%; Score 40.8; DB 5; Length 1056;
Best Local Similarity 56.8%; Pred. No. 1.7;
Matches 75; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

Qy 585 GGATATTTTAGAGCTCGAAGACACTCCTTGTATCATCTGTCGCGTGAAGCTGGGTGA 644
Db 639 GGTAGTTAGGCTCTAGATGAGGATGATTGCTATTCATCTCTAGTGGGAAATGGATGA 580
Qy 645 GAAGATCAGTACCATGATTTTCTATTTGATCGCGCGCAATCAAGGTATCTCATTTGT 704
Db 579 GAAGCGCCAGATTTTCGGATTTGTTGTTGTTAGTCTTATTCACACCCCTAAGGCAGT 520
Qy 705 ACAGGAAATGGC 716
Db 519 AGAAGAGATGC 508

RESULT 15
LOCUS AY294501/c 1064 bp DNA linear VRT 28-FEB-2004
DEFINITION Aerodramus vanikorensis specimen-voucher DHC01 NADH dehydrogenase
subunit 2 (ND2) gene, partial cds; mitochondrial.
ACCESSION AY294501
VERSION AY294501.1 GI:37812874
KEYWORDS
SOURCE Mitochondrion Aerodramus vanikorensis
ORGANISM Aerodramus vanikorensis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Apodiformes; Apodidae; Aerodramus.
1 (bases 1 to 1064)
PRICE,J.J., Johnson,K.P. and Clayton,D.H.
The evolution of echolocation in swiftlets
J. Avian Biol. 35 (2), 135-143 (2004)

REFERENCE
AUTHORS Price,J.J., Johnson,K.P. and Clayton,D.H.
TITLE The evolution of echolocation in swiftlets
JOURNAL J. Avian Biol. 35 (2), 135-143 (2004)
FEATURES
source Location/Qualifiers
1..1064
/organism="Aerodramus vanikorensis"
/mol_type="mitochondrial"

ORIGIN
Query Match 5.4%; Score 40.8; DB 5; Length 1056;
Best Local Similarity 56.8%; Pred. No. 1.7;
Matches 75; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

Qy 585 GGATATTTTAGAGCTCGAAGACACTCCTTGTATCATCTGTCGCGTGAAGCTGGGTGA 644
Db 639 GGTAGTTAGGCTCTAGATGAGGATGATTGCTATTCATCTCTAGTGGGAAATGGATGA 580
Qy 645 GAAGATCAGTACCATGATTTTCTATTTGATCGCGCGCAATCAAGGTATCTCATTTGT 704
Db 579 GAAGCGCCAGATTTTCGGATTTGTTGTTGTTAGTCTTATTCACACCCCTAAGGCAGT 520
Qy 705 ACAGGAAATGGC 716
Db 519 AGAAGAGATGC 508
```

```
REFERENCE
AUTHORS Price,J.J., Johnson,K.P. and Clayton,D.H.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2003) Center for Biodiversity, Illinois Natural
History Survey, 607 East Peabody Drive, Champaign, IL 61820, USA
FEATURES
source Location/Qualifiers
1..1056
/organism="Aerodramus terraereginae"
/mol_type="mitochondrial"
/mol_type="genomic DNA"
/specimen_voucher="DHC20"
/db_xref="taxon:46487"
37..>1056
/gene="ND2"
37..>1056
/gene="ND2"
/codon_start=1
/transl_table=2
/product="NADH dehydrogenase subunit 2"
/protein_id="AAR04261.1"
/db_xref="GI:37812897"
/translation="MNPYAKLIFTMSLLGTMTISSNHWMAWTGIEINTLAIIBLI
AKPHPRATIEAAIKYVLQAAATLLLFSSINAWSTGQMDITQLANPISCLILLTAI
AMKLGIVPFHPFPEVLQSSLTAMLSVMKFPPLTILFMTSPSLDPTLTTMAIS
STALGGMGLNTOIRKILAFSSISHLGWMAILIYSPKLTLLTFMYCLMTVTFLIT
LNTTKLSTWMTSWAKAPMLNTLMTLLSLAGLPPLTGPLPKLLIQELTKQEMT
TAATAIALISLGLFYLRLAYSTITLPPNPNNHMKQHIYKPTNTLISVLTSLSVL
LILPLSP"

ORIGIN
Query Match 5.4%; Score 40.8; DB 5; Length 1056;
Best Local Similarity 56.8%; Pred. No. 1.7;
Matches 75; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

Qy 585 GGATATTTTAGAGCTCGAAGACACTCCTTGTATCATCTGTCGCGTGAAGCTGGGTGA 644
Db 639 GGTAGTTAGGCTCTAGATGAGGATGATTGCTATTCATCTCTAGTGGGAAATGGATGA 580
Qy 645 GAAGATCAGTACCATGATTTTCTATTTGATCGCGCGCAATCAAGGTATCTCATTTGT 704
Db 579 GAAGCGCCAGATTTTCGGATTTGTTGTTGTTAGTCTTATTCACACCCCTAAGGCAGT 520
Qy 705 ACAGGAAATGGC 716
Db 519 AGAAGAGATGC 508

RESULT 15
LOCUS AY294501/c 1064 bp DNA linear VRT 28-FEB-2004
DEFINITION Aerodramus vanikorensis specimen-voucher DHC01 NADH dehydrogenase
subunit 2 (ND2) gene, partial cds; mitochondrial.
ACCESSION AY294501
VERSION AY294501.1 GI:37812874
KEYWORDS
SOURCE Mitochondrion Aerodramus vanikorensis
ORGANISM Aerodramus vanikorensis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Apodiformes; Apodidae; Aerodramus.
1 (bases 1 to 1064)
PRICE,J.J., Johnson,K.P. and Clayton,D.H.
The evolution of echolocation in swiftlets
J. Avian Biol. 35 (2), 135-143 (2004)

REFERENCE
AUTHORS Price,J.J., Johnson,K.P. and Clayton,D.H.
TITLE The evolution of echolocation in swiftlets
JOURNAL J. Avian Biol. 35 (2), 135-143 (2004)
FEATURES
source Location/Qualifiers
1..1064
/organism="Aerodramus vanikorensis"
/mol_type="mitochondrial"

ORIGIN
Query Match 5.4%; Score 40.8; DB 5; Length 1056;
Best Local Similarity 56.8%; Pred. No. 1.7;
Matches 75; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

Qy 585 GGATATTTTAGAGCTCGAAGACACTCCTTGTATCATCTGTCGCGTGAAGCTGGGTGA 644
Db 639 GGTAGTTAGGCTCTAGATGAGGATGATTGCTATTCATCTCTAGTGGGAAATGGATGA 580
Qy 645 GAAGATCAGTACCATGATTTTCTATTTGATCGCGCGCAATCAAGGTATCTCATTTGT 704
Db 579 GAAGCGCCAGATTTTCGGATTTGTTGTTGTTAGTCTTATTCACACCCCTAAGGCAGT 520
Qy 705 ACAGGAAATGGC 716
Db 519 AGAAGAGATGC 508
```

```
/specimen_voucher="DHC01"
/db_xref="taxon:243317"
25...>1064
/gene="ND2"
25...>1064
/gene="ND2"
/codon_start=1
/transl_table=2
/product="NADH dehydrogenase subunit 2"
/protein_id="AAR04251.1"
/db_xref="GI:37812875"
/translation="MNPYAKLIPTMSLLGCTTWTISSNHVTAAGLEINTLAIPLI
AKPHPRIRIAAIKYFLVQAAASTLLLFSSINAWSTGQWDITQLNPNVSCILLITAI
AMKLGVPFFHFPVEVLQGSPLFTAMLLSTVMKPPPTILFMTSPSLNPTLLATMAIS
STALGWMGLNQTKIRKILAFSSISHLGWMAIILIYSPKLTLLTFMYCMLTITVFLT
LNTTKTLKSTMTSWTKTPMLSTLMMTLISLAGLPPLTGFLPKWLIIOELTKQEMT
AAATVISILSLGLFFYLRLAYYSTITLPPNPTNHMKQWHIYKPTNTLISILTSLSIL
LLPLSPMLSSST"

ORIGIN
Query Match      5.4%; Score 40.8; DB 5; Length 1064;
Best Local Similarity 56.8%; Pred. No. 1.7;
Matches 75; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

Qy 585 GGATATTTTAGAGTGAAGAGACACTCCTTGCTTATCATCGTCCGTGGAGCTGGGTGA 644
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
627 GGTCACTTTGGGGCTGTAATGAGGATGCTTATTCATCCGAGGTGGGAGATGGATGA 568
Qy 645 GAAGATCAGTAACCATCAATTTCTATTGTCGCGCGCAATCAAGGGTATCTCATTTG 704
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
567 GAAGGCTAGGATCTTCGGATTGTGTGTTGTTAGTCCCATTCACCTCTAAGCGCGT 508
Qy 705 ACAGGAAATGGC 716
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
507 AGAAGAAATTGC 496
```

Search completed: December 10, 2004, 16:12:29
Job time : 3355.71 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2004, 13:25:54 ; Search time 18.7485 Seconds
(without alignments)
1288.126 Million cell updates/sec

Title: US-10-627-124-8
Perfect score: 1289
Sequence: 1 MNMTGVTERRSVHFAIA.....YLIVQEMAQLSEKRITSEG 251
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1273	98.8	248	2 A70044	conserved hypotet
2	215	16.7	289	2 G70730	hypothetical prote
3	102.5	8.0	608	2 T02299	hypothetical prote
4	98.5	7.6	332	2 H00174	hypothetical prote
5	94.5	7.3	3418	1 G02334	chromosome segrega
6	93	7.2	1156	2 E69444	breast cancer tumo
7	92	7.1	393	2 A71154	hypothetical prote
8	90	7.0	198	2 T01340	hypothetical prote
9	90	7.0	1198	2 D96723	hypothetical prote
10	89.5	6.9	134	2 S14947	2S albumin - Brazi
11	89.5	6.9	656	2 T52064	dnad-like protein
12	89.5	6.9	898	2 E96659	hypothetical prote
13	89	6.9	420	2 G90505	phosphoribosylamin
14	87.5	6.8	977	2 E86349	hypothetical prote
15	87	6.7	355	2 AE3008	monooxygenase limp
16	87	6.7	355	2 H98275	mtaG protein (Af18
17	87	6.7	365	1 A26522	3-isopropylmalate
18	87	6.7	951	2 G82965	conserved hypotet
19	86.5	6.7	255	2 S66068	conserved hypotet
20	85	6.6	420	2 D97878	phosphoribosylamin
21	85	6.6	716	2 G64816	probable ATP-depen
22	85	6.6	791	2 H72552	hypothetical prote
23	85	6.6	1172	2 C70619	probable lyxX prot
24	84.5	6.6	234	2 F96620	hypothetical prote
25	84	6.5	299	2 G90677	cyn operon positiv
26	84	6.5	299	2 G85528	cyn operon positiv
27	84	6.5	716	2 E90738	probably ATP-depen
28	84	6.5	716	2 G85588	probably ATP-depen
29	84	6.5	818	2 S62790	mismatch DNA recog

ALIGNMENTS

RESULT 1
A70044
conserved hypothetical protein ymc - Bacillus subtilis

C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004

C;Accession: A70044

R;Kunst, F.; Ogaawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter...

A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A;Accession: A70044

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-248 <KUN>

A;Cross-references: UNIPROT:O34351; GB:Z99121; GB:Z99122; GB:AL009126; NID:G2636029; PIDN

A;Experimental source: strain 168

C;Genetics:

30 83.5 6.5 529 2 F71547
31 83.5 6.5 1170 2 A72287
32 83 6.4 804 1 A69309
33 82.5 6.4 394 2 A70469
34 82 6.4 360 2 A49600
35 82 6.4 499 2 F72768
36 82 6.4 864 2 H83386
37 81.5 6.3 301 2 C90264
38 81.5 6.3 325 2 F68811
39 81 6.3 592 2 G69252
40 81 6.3 663 2 I46611
41 81 6.3 1018 2 AH0293
42 81 6.3 1302 2 T20767
43 81 6.3 1767 2 T20766
44 80.5 6.2 624 2 T04843
45 80 6.2 299 2 A41900

mutated
not a copy

db 241 EKRITSEG 248
|||||

RESULT 2

G70730
 Hypothetical protein Rv2275 - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jun-2004
 C:Accession: G70730
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Cole, S.T.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
 A:Authors: Seares, R.; Suleston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98255987; PMID:9634230
 A:Accession: G70730
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-289 <COL>
 A:Cross-references: UNIPROT:Q50689; GB:277163; GB:AL123456; NID:g3261610; PIDN:CAB00960
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: Rv2275

Query Match	16.7%	Score 215;	DB 2;	Length 289;
Best Local Similarity	27.3%;	Pred. No. 7.2e-11;		
Matches	62;	Conservative	34;	Mismatches 115;
				Indels 16;
				Gaps 4;
Qy	16	FIAEALTENCREIFERRHHVLVGISPNRSFSEDIYRLIGWAKAQKFSVLLAGHEAA	75	
Db	64	FLVRPTQQCCIHTGEDHAVIGSPGNSVFSRORLDRDLGLWGLTNDFRDVDFVTDVHVA	123	
Qy	76	NLLLEALTGPGRKAERKVRKESVNRNRFAERALVAHGDDPK----	AHTEFSDFIDNKAYQL	131
Db	124	ESYEALGDSALERRKAVKQIRGVR--AKITTTWNELOPAGARLCVRPMSEFQSNAYRE	181	
Qy	132	LRQEVHAFPEQPHFRACLDMSREAL----	IGRARGVSLMEEVSESDMLNLAVEYVIAEL	188
Db	182	LHADLLTRLKDDDEDLRAVCODLVRFRFSTKVGPFQATATQEQV-----	CMDYICABA	234
Qy	189	PFTGAPDIILEVBETLLAYHRPWLKGIKSIHNHFSICWRPNQGYLIV	235	
Db	235	PIFLDTALLICVPSNLCYHOSIPIARMLVARGSGSLRASRNOCHAVI	281	

RESULT 3

T02299
hypothetical protein F23858.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 09-Jul-2004
C:Accession: T02299; T46353
R:Iamerdin, J.E.; McCready, P.M.; Skowronski, E.; Adamson, A.W.; Burkhardt-Schultz, K.; Frankhe
; Christensen, M.; Georgescu, A.; Avila, J.; Liu, S.; Attix, C.; Andreise, T.; Frankhe
submitted to the EMBL Data Library, March 1998
A:Authors: Duarte, S.; Lucas, S.; Bruce, R.; Thomas, P.; Quan, G.; Kronmiller, B.; Arell
A:Description: Sequence analysis of a human P1 clone containing the XRCC9 DNA repair gen
A:Reference number: Z14637
A:Accession: T02299
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-608 <I>AM>
A:Cross-references: UNIPROT:O60378; EMBL:AC004475; NID:g2988396; PIDN:AAC08052.1; PID:g2
R:Koehrer, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23037
A:Accession: T46353
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 217-608 <AAA>
A:Cross-references: EMBL:AL137286
A:Experimental source: adult testis; clone DKFZp434e2216

C;Genetics:
A;Map position: 19
A;Introns: 58/1; 8
A;Note: DKFZp434E2

Query Match	8.0%;	Score 102.5;	DB 2;	Length 608;
Best Local Similarity	26.3%;	Pred. No. 0.73;		
Matches	45;	Conservative 24;	Mismatches 71;	Indels 31; Gaps 8;
Qy	84	PRGKAERKVKYKVRNRRFAERALVAHGDDPKAIHTFSDPDKNAQYLLRQVEVHAFFEQ	143	
Db	146	PEGAETRKVKIEKLAR-----FVAGGPELEKVMEDYKONPAFLHDKNSRFL--	195	
Qy	144	PHFRHACLDWSREALICARGVSLMMEEVSDMLNAVYVIAELPFFI--GAPDILEVE	201	
Db	196	-YRKVKVAIRKEAQKSAQKVSPP-EDEEVKLA-----EKLARFIADGGP---EVE	245	
Qy	202	ETLLAYHRPWLKGEKITSNHFSTCMRPN-QGYLIVQEMQMSEKRITSG	251	
Db	246	TIALONNR-----ENQAFSLYFENSGYKYKVKLEEFKPKASSTFG	288	

RESULT 4

H90174
 Hypothetical protein SSO0320 [imported] - Sulfolobus solfataricus
 C:Species: Sulfolobus solfataricus
 C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
 C:Accession: H90174
 R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.;
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.;
 arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
 submitted to GenBank, April 2001
 A:Description: Sulfolobus solfataricus complete genome.
 A:Reference number: A99139
 A:Accession: H90174
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-392 <KUR>
 A:Cross-references: UNIPROT:Q980H7; GB:AE006641; NID:g13813463; PIDN:AAK40655.1
 C:Genetics:
 A:Gene: SSO0320

[illegible]

RESULTS

breast cancer tumor suppressor BRCA2 - human
N/Alternate names: breast cancer susceptibility protein BRCA2
C/Species: Homo sapiens (man)
C/Date: 21-Dec-1996 #sequence revision 06-Jun-1997 #text change 09-Jul-2004
G02334

```

Query Match          7.2%; Score 93; DB 2; Length 1156;
Best Local Similarity 19.2%; Pred. No. 11;
Matches 39; Conservative 47; Mismatches 59; Indels 58; Gaps 9;

QY      1  MNEMTGWVTERSS-----VHFIAEALTENCREIFERRHHVLVGISPNSR 45
Db      707  VDRLTGWISLNRNRIILDEKIRTESGRIBELREKISQKSR-----KENYISLKDYSK 762

QY      46  FSDYIYRLIGWAKAQKSVSVLLAGHEAANLLEALGTPRGKAEKVKRKEVSRNRRFAER 105
Db      763  LAE--MEEAIGEALAEABIEETIERMLRGSEVPKIVVEILD-----KIKESHQRNREI-- 809

QY      106  ALVAHGDDPKAIHTFSDFDINK--AYQLLRQVEVHAFFPEOPHFHACLD---MSREAIL 159
Db      810  -LIS-----LEKKTESIEFFRQREQLSSMQEKQVYLDKIDRIDEIRRIIEE 854

QY      160  GRARGVSLMMEEVSEDMNLAVE 182
Db      855  GKAR-----VEEINSELEELRKE 872

RESULT 7
A71154
hypothetical protein PH0432 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 09-Jul-2004
C:Accession: A71154
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, N.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, M.
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic bacterium, Pyrococcus horikoshii
A:Reference number: A71000; MUID:98344137; PMID:9679194
A:Accession: A71154
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-393 <RAW>
A:Cross-references: UNIPROT:O58169; GB:AP000002; NID:G3236129; PIDN:BAA29518.1;
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by A71154
C:Genetics:
A:Gene: PH0432

Query Match          7.1%; Score 92; DB 2; Length 1393;
Best Local Similarity 21.2%; Pred. No. 3.4;
Matches 56; Conservative 40; Mismatches 92; Indels 76; Gaps 11;

QY      30  ERRRHVLVGISPNSRF--SEDYIYRLIGWAKAQKSVSVLLAGHEAANLLEALGTPRGK 87
Db      5  EKPIAIIGLRGIPSKYGGTETFEVELTSRLKDQAFK---FYVMHEDTKFFE----- 52

QY      88  AERKVRKEVSRNRRFAERLVAHGDDPKAIHTFSDFDINKAYQLLRQVEVHAFFPEOPH 145
Db      53  -----DEYNRIIRVHSPAIESKTSIPSINDFT---NTAYMLANHEKDIELPYFLGP- 101

QY      146  FRHACLDMSREALIGRARGVSLMMEEVSEDMNLMA-----VE 182
Db      102  -----DSSLAAIILRLGKKVILNDPGVWRRLIKRESYFVPFLFYIYFATIIYMFME 155

QY      183  YVIAELPFFTGAPDILEVETTLAYHRPWKL-----GEKISNHFPSI-----CM 226
Db      156  YLSCKLPDIWA--DSLGIKEHLEKHKRKPRAWVITYGARELISSEFSVEBEKEILSRFL 214

QY      227  RPNQGYLIYQEMAQMLSEKHSITSE 250
Db      215  EPLGYLTV---ARIVANNIHME 235

RESULT 8
T01340
hypothetical protein F6N15.18 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004

```


Qy 145 ---HFRHACLDGREA1--IGRARGVSLMWE-EVSDMLNLAVEYVIAELPPFIGAPDIL 198
Db 475 GYIVROAARELGKSA1ALMQLQEGMKMBESDNKEDQIMKSPBEKKDAML5YLWKINVV 534
Qy 199 EVESTL 204
Db 535 DIESTL 540

RESULT 12
E96659
hypothetical protein F9N12.3 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: E96659
R:Theologias, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Huizar, L.
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: E96659
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-898 <STO>
A:Cross-references: UNIPROT:Q9C8T9; GB:AE005173; NID:g6997185; PIDN:AAF34847.1; GSPDB:GN
C:Genetics:
A:Map position: 1
C:Superfamily: disease resistance protein RPS2; leucine-rich alpha-2-glycoprotein repeat

Query Match 6.9%; Score 89.5; DB 2; Length 898;
Best Local Similarity 20.1%; Pred. No. 15;
Matches 57; Conservative 49; Mismatches 84; Indels 93; Gaps 13;
Qy 1 MNEMGTWYTER---RSVHFIAEALTENCREFERRHVLVIGSPNSRFS--EDYIYR 53
Db 333 IRELSRVVAKCCGLPALNVSE--TWCKRTVQEWHRHAIYVLNSYAAKPSGMDKILP 390
Qy 54 LIGWAKAQFQSVSVLLAGHEAANLLEALGTGPRGKAERKVRKVSRRNRRFAERALVAHGGD 113
Db 391 LLKYSYDSLKGEDVQMC-----LLYCALPFE---DAKIRKENLIEWICEEIIIDGSEGI 441
Qy 114 PKAIHTSPDFIDNKAYQLLRQVEHAFPEQPHFRHACLDMSREAIIGRARGVSLMEEVS 173
Db 442 DKA-----ENQGYE-----IIGSLVRASILLMEEVE 466
Qy 174 EDMLNL-AVEYVIAELPPFI-----GAPDILEVEETLLAYHRPWKLGE 215
Db 467 LGGANIVCLHDVREMAWLTASDLGKQNEAFIVRASVGLREILKVEN-----MNVVR 518
Qy 216 KISNHEPSICWRPN-----QGYLIQVQEAQMLSE-----KRITSE 250
Db 519 RMS-----LMKNNIAHLDRLCDMELTLLQLQSTHLEKISSE 555

RESULT 13
G95005
phosphoribosylamine-glycine ligase [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004
C:Accession: G95005
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzaple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: G95005
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-420 <KUR>
A:Cross-references: UNIPROT:Q97T98; GB:AE005672; PIDN:AAK74240.1; PID:g14971515; GSPDB:GN
C:Genetics:
A:Map position: 1
C:Superfamily: phosphoribosylamine-glycine ligase; phosphoribosylamine-glycine ligase hom

Query Match 6.9%; Score 89; DB 2; Length 420;
Best Local Similarity 23.2%; Pred. No. 6.6;
Matches 57; Conservative 34; Mismatches 89; Indels 66; Gaps 10;
Qy 15 HFTAEALTE--NCREIFFERRH---VLVIGSPNSRFSYDIYRLIGWAKAQFQSVSL 69
Db 13 HATAKKLESKQVEKVFVAPGNDGMTLDGLVNLISSEH--YKLIDFAKT--NDVAWTF 68
Qy 70 AGHE---ANWL-----LEALGTGPRGKAERKVRKVSRRNRRFAERALVAHGGDPKAIH 118
Db 69 IGPDALLAAGIVDDFNQAGLKAFGPTRAAAALEWSKD-----FAKEIMVKYGVPTATYG 122
Qy 119 TFSDFIDNKAYQLLRQVEHAFPEQPHFRHACLDMSREAIIGRARGVSL---MMEEVSE 174
Db 123 TFSDFBEAKAY-----TEKHGAPIVVRADGLAGLKGKGVVVAETVE 161
Qy 175 DMLNLAVEYVIAELPPFI GAPDILEVEETLLAYHRPWKLGEKISNHEFSICMRPNQGYLI 234
Db 162 QAVEAAHEMLLDNKGDSGARVIE-----EFLEGEFEFSLFAFVNGDKFY 206
Qy 235 VQEMAQ 240
Db 207 IMPTAQ 212

RESULT 14
E86349
hypothetical protein F8K7.6 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: E86349
R:Theologias, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Huizar, L.
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: E86349
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-977 <STO>
A:Cross-references: UNIPROT:Q9XI14; GB:AE005172; NID:g5263315; PIDN:AAD41417.1; GSPDB:GN
C:Genetics:
A:Map position: 1
C:Superfamily: preprotein translocase secA

Query Match 6.8%; Score 87.5; DB 2; Length 977;
Best Local Similarity 22.8%; Pred. No. 25;
Matches 62; Conservative 50; Mismatches 85; Indels 75; Gaps 19;
Qy 1 MNEMGTWYTER---SVHFIAEALTENCREFERRHVLVIGSPNSRFSYDIY--- 52
Db 385 INELTGRVEDKRWSEGVHQAVERA---KEGLEIQADSIWVAQITQSLFK---LYPKL 436
Qy 53 -RLIHWAKAQFK-----SVSVLLAGHEAANL-----LEALGTGPRGKAERKVRKVSRRN 100
Db 437 SGMTGTAKTEBEKFLKMFQIPVIEVPTNLNIRIDLPQAFATARGKWEH--VRREV--ED 493

```
QY 101 RPAERALVAHGDPKAIHTSFIDNKAY--QLLRQVEHAFPEQPHF-----RHACLD 152
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 494 MFCQ-----GRPVLVGTTS--VENSEYVSELLKE-----WGIPHNVLNARPKYAARE 538
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 153 MSREAIIGRARGVSLMMEEVSEDMNLNAVYVIAELPFFIGAPDILE---VEETLLAYHR 209
      | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 539 ADFIAQAQRKYAIT-----ISTNAGRGTDIILG-----GNPKMLAREIIEDSILSYLT 587
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 210 PWKLGKISHNHFPSICMRPNQGYLIVQEMAQM 241
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 588 SEVLADNIDDELS---QKN-----LINEQSEM 612
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 15
AE3008
monooxygenase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:Accession: AE3008
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
; Karp, P.; Gillet, W.; Grant, C.; Guentherer, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AE3008
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-355 <KUR>
A:Cross-references: UNIPROT:Q8U9Q5; GB:AE008689; PIDN:AAL44483.1; PID:g17742088; GSPDB:C
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu3671
A:Map position: linear chromosome
C:Superfamily: alkanal monooxygenase [FMN-linked] (bacterial luciferase)

Query Match 6.7%; Score 87; DB 2; Length 355;
Best Local Similarity 23.5%; Pred. No. 7.9;
Matches 36; Conservative 28; Mismatches 63; Indels 26; Gaps 7;

QY 53 RLIGWAKAQFYSVLLAGHEAANLLEALGTPRGKAERKYRKEVSRNRFPARALVAHGG 112
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 166 RIFPFWACQ-STETFIAGRRGLNLTAL---LGTTESLAPKIAAYKRERK-----HGH 217
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 113 DPKA-----IHTFSDPIDNKAYQLLRQVEHAFPEQPHFRHACLDMSREAIIGRARGVS 166
      | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 218 DPAAGTVSMNVHTFLGDDTCK-----VKSNSVKPFGDYLRTHYHLLG-----LGRSMGVD 268
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 167 LMMEEVSEDMNLNAVYVIAEL---PFFIGAPD 196
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 269 ITLDNFSDDDLSDLSLIEFGIEGFMKGRSLIGTPE 301
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
```

Search completed: December 10, 2004, 13:45:52
Job time : 21.7485 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2004, 12:47:24 ; Search time 96.4207 Seconds
(without alignments)
1497.801 Million cell updates/sec

Title: US-10-627-124-8

Perfect score: 1289

Sequence: 1 MNEMTGVNTERSVMHFAEA.....YLIVQEMAQWLSEKRITSESG 251

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 02.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1273	98.8	248	2	Q34351
2	400.5	31.1	234	2	Q7N9M5
3	249	19.3	239	2	Q8GED7
4	215	16.7	289	2	Q7TVV7
5	215	16.7	289	2	Q50688
6	102.5	8.0	608	2	Q60378
7	102.5	8.0	641	2	Q8WWT4
8	102.5	8.0	643	2	Q6P3X9
9	102.5	8.0	643	2	AH63784
10	102.5	8.0	645	2	Q8IWX8
11	100.5	7.8	700	2	Q8XUM4
12	100	7.8	1093	2	Q82527
13	98.5	7.6	392	2	Q980H7
14	97	7.5	864	2	Q75130
15	94.5	7.3	3418	1	BRC2_HUMAN
16	93	7.2	1156	2	Q28714
17	92	7.1	393	2	Q58169
18	91	7.1	652	2	Q9F531
19	91	7.1	751	2	Q86210
20	90.5	7.0	360	2	Q95XC8
21	90.5	7.0	366	2	Q6L118
22	90	7.0	916	2	Q94E26
23	90	7.0	1198	2	Q04538
24	89.5	6.9	382	2	Q7NTG4
25	89.5	6.9	556	2	Q9SH21
26	89.5	6.9	656	2	Q9XEM8
27	89.5	6.9	898	1	DR19_ARATH
28	89	6.9	420	1	PUR2_STRPN
29	88.5	6.9	399	2	Q8CV34
30	88	6.8	285	2	Q9F5V9
31	88	6.8	643	2	Q8CH02

RESULT 1

ID	Q34351	PRELIMINARY;	PRT;	248 AA.
AC	Q34351; Q795E6;			
DT	01-JAN-1998 (TrEMBLrel. 05, Created)			
DT	01-JAN-1998 (TrEMBLrel. 05, Last sequence update)			
DT	01-OCT-2004 (TrEMBLrel. 28, Last annotation update)			
DE	YvmC.			
GN	Names=yvmC; OrderedLocusNames=BSU35070;			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
OX	NCBI_TaxID=1423;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=98230327; PubMed=9570401;			
RA	Reizer J., Holscher C., Titgemeyer F., Rivolta C., Rabus R.,			
RA	Stulke J., Karamata D., Saier M.H. Jr., Hillen W.;			
RT	"A novel protein kinase that controls carbon catabolite repression in			
RT	bacteria.";			
RL	Mol. Microbiol. 27:1157-1169 (1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Lazarevic V., Soldo B., Rivolta C., Reynolds S., Mael C.,			
RL	Karamata D.;			
RP	Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RX	MEDLINE=98044033; PubMed=9384377; DOI=10.1038/36786;			
RA	Kunst F., Ogasawara N., Moser I., Albertini A.M., Alloni G.,			
RA	Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,			
RA	Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,			
RA	Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,			
RA	Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,			
RA	Denicot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,			
RA	Entian K.-D., Errington J., Fabret C., Ferrari E., Foulger D.,			
RA	Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,			
RA	Ghim S.-Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,			
RA	Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,			
RA	Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M.,			
RA	Jones L.-W., Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M.,			
RA	Klein C., Kobayashi Y., Koetter P., Koningsstein G., Krogh S.,			
RA	Kumano M., Kurita K., Lapidus A., Lardinois S., Lauber J.,			
RA	Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Mael C.,			
RA	Medigue C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S.,			
RA	Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B.,			
RA	Park S.H., Parro V., Pohl T.M., Portetelle D., Porwollik S.,			
RA	Prescott A.M., Presecan E., Pujic P., Purnelle B., Rapoport G.,			
RA	Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B.,			
RA	Rose M., Sadaie Y., Sato T., Scanlan A., Schleich S., Schroeter R.,			
RA	Scoffone F., Sekiguchi J., Sekowska A., Seror S.J., Serror P.,			
RA	Shin B.S., Soldo B., Sorokin A., Tacconi E., Takagi T., Takahashi H.,			
RA	Takenaru K., Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P.,			
RA	Tognoni A., Tosato V., Uchiyama S., Vandenbol M., Vannier F.,			

ALIGNMENTS

32	88	6.8	646	2	Q8R094
33	88	6.8	646	2	AH27188
34	88	6.8	1058	2	Q74260
35	88	6.8	1058	2	AA54836
36	87.5	6.8	292	2	Q8A7C1
37	87.5	6.8	895	2	Q89JF5
38	87.5	6.8	975	2	Q8ZWR5
39	87.5	6.8	977	2	Q9X114
40	87	6.7	355	2	Q7C116
41	87	6.7	355	2	Q8U9Q5
42	87	6.7	365	1	LEU3_BACSU
43	87	6.7	501	2	Q7ZXV7
44	87	6.7	647	2	Q7Y145
45	87	6.7	691	2	Q6WL05

Q8R094 mus musculu
Aah27188 mus muscu
Q74260 ashbya goss
Aas54836 ashbya go
Q8A7C1 bacteroides
Q89JF5 bradyrhizob
Q8ZWR5 pyrobaculum
Q9X114 arabidopsis
Q7C116 agrobacteri
Q8U9Q5 agrobacteri
P05645 bacillus su
Q7ZXV7 xenopus lae
Q7Y145 oryza sativ
Q6WL05 panulirus a

RA Vassarotti A., Viari A., Wambutt R., Wedler E., Wedler H.,
RA Weitneger T., Winters P., Wipat A., Yamamoto H., Yanane K.,
RA Yasumoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumstein E.,
RA Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.";
RL Nature 390:249-256(1997).
DR ENBL; AF017113; AAC67279.1; --
DR ENBL; Z99121; CAB15512.1; --
DR PIR; A70044; A70044.
DR InterPro; IPR006162; Ppantne S.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 248 AA; 28506 MW; 17DF979833A0B1B8 CRC64;

Query Match 98.8%; Score 1273; DB 2; Length 248;
Best Local Similarity 100.0%; Pred. No. 5.3e-96;
Matches 248; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 MTGWTERRSVHFIATALTENCREIFERRRHVLVGISPFNSRSEDYIYRLIGWAKAQPK 63
DB 1 MTGWTERRSVHFIATALTENCREIFERRRHVLVGISPFNSRSEDYIYRLIGWAKAQPK 60
QY 64 SVSVLLAGHAANLLEALGTPRGKAERKVRKEVSRNRFAERLVAHGGDPKAIHTFSDF 123
DB 61 SVSVLLAGHAANLLEALGTPRGKAERKVRKEVSRNRFAERLVAHGGDPKAIHTFSDF 120
QY 124 IDNKAYQLRQVEHAFPEQPHFRHACLDMSREAIIGRARGVSLMMEVSEDMNLVAYEY 193
DB 121 IDNKAYQLRQVEHAFPEQPHFRHACLDMSREAIIGRARGVSLMMEVSEDMNLVAYEY 180
QY 184 VIAELPFFIGAPDILEVEETLLAYHRPWKLGEKISNHEFSICMRPNQGYLIVQEMAQMLS 243
DB 181 VIAELPFFIGAPDILEVEETLLAYHRPWKLGEKISNHEFSICMRPNQGYLIVQEMAQMLS 240
QY 244 EKRTISEG 251
DB 241 EKRTISEG 248

RESULT 2
Q7N9M5 PRELIMINARY; PRT; 234 AA.
AC Q7N9M5
DT 01-MAR-2004 (TREMBLrel. 26, Created)
DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Similar to unknown protein.
GN OrderedLocusNames=plu0297;
OS Photorhabdus luminescens (subsp. laumondii).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=141679;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TT01;
RX MEDLINE=22957627; PubMed=14528314;
RA Duchaud E., Rusniok C., Frangoul L., Buchrieser C., Givaudan A.,
RA Taouit S., Bocs S., Boursaux-Eude C., Chandler M., Charles J.-F.,
RA Dassa E., Derose R., Derzelle S., Freysinet G., Gaudriault S.,
RA Medigue C., Lanois A., Powell K., Siquier P., Vincent R., Wingate V.,
RA Zouine M., Glaser P., Boenare N., Danchin A., Kunst F.;
RT "The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*.";
RL Nat. Biotechnol. 21:1307-1313(2003).
DR ENBL; BX571859; CAB12592.1; --
DR Photolyst; plu0297; --
KW Complete proteome.
SQ SEQUENCE 234 AA; 26893 MW; B9291248C7B7E8B6 CRC64;

Query Match 31.1%; Score 400.5; DB 2; Length 234;
Best Local Similarity 37.6%; Pred. No. 1.2e-24;
Matches 86; Conservative 43; Mismatches 97; Indels 3; Gaps 2;

QY 13 SVHFTAEALTENCREIFERRRHVLVGISPFNSRSEDYIYRLIGWAKAQPKSVLL-AG 71
DB 6 SPSFTVOGETSCDQIIQGDHALIGISPFNSRFSKDYVDLIQMSHFRQVDILLPCE 65
QY 72 HPAANLLEALGTPRGKAERKVRKEVSRNRFAERLVAHGGDPKAIHT--FSDFINKAY 129
DB 66 REASRLLVASGIDNVKAIKKTHREIRRLNLDYVISTATLKSQRVTFQSFSLNHDY 125
QY 130 QLLRQVEHAFPEQPHFRHACLDMSREAIIGRARGVSLMMEVSEDMNLVAYEYVIAELP 189
DB 126 QSLKTQVENAFNESESFKKSCLDMSFQAIIKGLKGTGQYFGQIDQLQVVKALPYIFAEIP 185
QY 190 FFIGAPDILEVEETLLAYHRPWKLGEKISNHEFSICMRPNQGYLIVQEM 238
DB 186 FYLNTPRLLGVKYSYLLYHRPMSIGKGLFNGSPYIQVADKQSGYITQL 234

RESULT 3
Q8GED7 PRELIMINARY; PRT; 239 AA.
AC Q8GED7
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE AlbC.
GN Name=albc;
OS Streptomyces noursei.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1971;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 11455;
RA Lautru S., Gondry M., Genet R., Pernodet J.-L.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR ENBL; AY129235; AAN07909.1; --
SQ SEQUENCE 239 AA; 26774 MW; CD368019F6EF87F3 CRC64;

Query Match 19.3%; Score 249; DB 2; Length 239;
Best Local Similarity 29.6%; Pred. No. 3e-12;
Matches 67; Conservative 39; Mismatches 110; Indels 10; Gaps 4;

QY 19 EALTENCREIFERRRHVLVGISPFNSRSEDYIYRLIGWAKAQPKSVLLAGHAANLL 78
DB 16 EILGDRSLIRQGEHALIGISAGNSYFSQKNTWMLQWAGQFFERTDVVYVDTHIDEML 75
QY 79 EALGTPRGKAERKVRKEVSRNRFAERLVAHGGDPK--AIHTFSDFINKAYQLLRQEV 136
DB 76 IADGRSAQAERSVKRTLDLRRRLRSLESVGDHAERFVRSLSELQETPEYRAVRERT 135
QY 137 EHAFPEQPHFRHACLDMSREAIIGR-ARGVSLMMEVSEDMNLVAYEYVIAELPFFIGAP 195
DB 136 DRAFEEDAEFACEDMVRVNMRFPGDVG-----ISAEHLRAGLNVLAEPADFADSP 190
QY 196 DILEVEETLLAYHRPWKLGEKISNHEFSICMRPNQGYLIV--QEMA 239
DB 191 GVFSVSSVLCVHIDTPIITAFLSRRETGFRAEGQAYVVVRPQELA 236

RESULT 4
Q7TYV7 PRELIMINARY; PRT; 289 AA.
AC Q7TYV7
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein Mb2298.
GN OrderedLocusNames=Mb2298;
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;

Query Match 8.0%; Score 102.5; DB 2; Length 608;

```
Best Local Similarity 26.3%; Pred. No. 8.4;
Matches 45; Conservative 24; Mismatches 71; Indels 31; Gaps 8;

QY 84 PRGAERKVRKVEVSRNRRFAERLVAHGDPKAIHTFSDFDINKAYQLLRQVEVHAFFEQ 143
Db 146 PEGAETRKVIEKLAR-----FVAEGGPELEKVMEDYKDNPAFLHDKNSREFL-- 195

QY 144 PHFRHACLDMSREAITGRARGVSLMMEVSEDMNLNLAVEVYIAELPFFI--GAPDILEVE 201
Db 196 -YRKVKVAIRKKAQKSAQSKVSPPE--DEEVKNLA-----EKLARFIADGGP---EVE 245

QY 202 ETLLAYHRPWKLGEKISNHEFSICMRPN-QGYLIVQEMAQMLSEKRIITSG 251
Db 246 TIALQNNR-----ENQAFSLYEPNSQGYKYRQKLEEFKAKASSTG 288

RESULT 7
Q8WWT4 PRELIMINARY; PRT; 641 AA.
AC Q8WWT4
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE RNA-binding protein splice variant a.
GN Name=RBP;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Gu Y., Nguyen C.-T.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY072917; AAL68961.1; -.
DR Genew; HGNC:18643; SF4.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0006396; P:RNA processing; IEA.
DR InterPro; IPR000467; G_patch.
DR InterPro; IPR000061; Surp.
DR Pfam; PF01585; G_patch; 1.
DR Pfam; PF01805; Surp; 2.
DR PROSITE; PS50174; G_PATCH; 1.
DR PROSITE; PS50128; SURP; 2.
SQ SEQUENCE 641 AA; 72011 MW; 60BCA0B5EB1A0D80 CRC64;

Query Match 8.0%; Score 102.5; DB 2; Length 641;
Best Local Similarity 26.3%; Pred. No. 8.9;
Matches 45; Conservative 24; Mismatches 71; Indels 31; Gaps 8;

QY 84 PRGAERKVRKVEVSRNRRFAERLVAHGDPKAIHTFSDFDINKAYQLLRQVEVHAFFEQ 143
Db 179 PEGAETRKVIEKLAR-----FVAEGGPELEKVMEDYKDNPAFLHDKNSREFL-- 228

QY 144 PHFRHACLDMSREAITGRARGVSLMMEVSEDMNLNLAVEVYIAELPFFI--GAPDILEVE 201
Db 229 -YRKVKVAIRKKAQKSAQSKVSPPE--DEEVKNLA-----EKLARFIADGGP---EVE 278

QY 202 ETLLAYHRPWKLGEKISNHEFSICMRPN-QGYLIVQEMAQMLSEKRIITSG 251
Db 279 TIALQNNR-----ENQAFSLYEPNSQGYKYRQKLEEFKAKASSTG 321

RESULT 8
Q6P3X9 PRELIMINARY; PRT; 643 AA.
AC Q6P3X9
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE SF4 protein (Fragment).
GN Name=SF4;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC063784; AAH63784.1; -.
DR InterPro; IPR000467; G_patch.
DR Pfam; PF01585; G_patch; 1.
DR Pfam; PF01805; Surp; 2.
DR SMART; SM00443; G_patch; 1.
DR SMART; SM00648; SWAP; 2.
DR PROSITE; PS50174; G_PATCH; 1.
DR PROSITE; PS50128; SURP; 2.
FT NON TER 1
SQ SEQUENCE 643 AA; 72252 MW; 4DE9A061B9C1C2D1 CRC64;

Query Match 8.0%; Score 102.5; DB 2; Length 643;
Best Local Similarity 26.3%; Pred. No. 8.9;
Matches 45; Conservative 24; Mismatches 71; Indels 31; Gaps 8;

QY 84 PRGAERKVRKVEVSRNRRFAERLVAHGDPKAIHTFSDFDINKAYQLLRQVEVHAFFEQ 143
Db 181 PEGAETRKVIEKLAR-----FVAEGGPELEKVMEDYKDNPAFLHDKNSREFL-- 230

QY 144 PHFRHACLDMSREAITGRARGVSLMMEVSEDMNLNLAVEVYIAELPFFI--GAPDILEVE 201
Db 231 -YRKVKVAIRKKAQKSAQSKVSPPE--DEEVKNLA-----EKLARFIADGGP---EVE 280

QY 202 ETLLAYHRPWKLGEKISNHEFSICMRPN-QGYLIVQEMAQMLSEKRIITSG 251
Db 281 TIALQNNR-----ENQAFSLYEPNSQGYKYRQKLEEFKAKASSTG 323

RESULT 9
AAH63784 PRELIMINARY; PRT; 643 AA.
AC AAH63784
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE SF4 protein (Fragment).
GN SF4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC063784; AAH63784.1; -.
DR InterPro; IPR000467; G_patch.
DR Pfam; PF01585; G_patch; 1.
DR Pfam; PF01805; Surp; 2.
DR SMART; SM00443; G_patch; 1.
DR SMART; SM00648; SWAP; 2.
DR PROSITE; PS50174; G_PATCH; 1.
DR PROSITE; PS50128; SURP; 2.
FT NON TER 1
SQ SEQUENCE 643 AA; 72252 MW; 4DE9A061B9C1C2D1 CRC64;

Query Match 8.0%; Score 102.5; DB 2; Length 643;
Best Local Similarity 26.3%; Pred. No. 8.9;
Matches 45; Conservative 24; Mismatches 71; Indels 31; Gaps 8;

QY 84 PRGAERKVRKVEVSRNRRFAERLVAHGDPKAIHTFSDFDINKAYQLLRQVEVHAFFEQ 143
Db 181 PEGAETRKVIEKLAR-----FVAEGGPELEKVMEDYKDNPAFLHDKNSREFL-- 230

QY 144 PHFRHACLDMSREAITGRARGVSLMMEVSEDMNLNLAVEVYIAELPFFI--GAPDILEVE 201
Db 231 -YRKVKVAIRKKAQKSAQSKVSPPE--DEEVKNLA-----EKLARFIADGGP---EVE 280

QY 202 ETLLAYHRPWKLGEKISNHEFSICMRPN-QGYLIVQEMAQMLSEKRIITSG 251
Db 281 TIALQNNR-----ENQAFSLYEPNSQGYKYRQKLEEFKAKASSTG 323

RESULT 9
AAH63784 PRELIMINARY; PRT; 643 AA.
AC AAH63784
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE SF4 protein (Fragment).
GN SF4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC063784; AAH63784.1; -.
DR InterPro; IPR000467; G_patch.
DR Pfam; PF01585; G_patch; 1.
DR Pfam; PF01805; Surp; 2.
DR SMART; SM00443; G_patch; 1.
DR SMART; SM00648; SWAP; 2.
DR PROSITE; PS50174; G_PATCH; 1.
DR PROSITE; PS50128; SURP; 2.
FT NON TER 1
SQ SEQUENCE 643 AA; 72252 MW; 4DE9A061B9C1C2D1 CRC64;
```



```
Db 233 HEAALRVLDLSTFAPOHPAALATARANVLKRVRLDEALAVARQAVVLAPRSAAEHALLA 292
QY 122 DFDNKAYQLLRQVEHAPQPHRACL--DMSREALIGARGVSLMMVEVSESDMLN 178
Db 293 -----MALQTLGTQDE-----ALPHFQOARLPQAVAEALVGR-----TLMEAGRDAAR 340
QY 179 LAVEVYIAELP-----FFIGAPDILEVEETLLAYHRPWKLGKISNHEFSI 224
Db 341 AAFDQALEQPGSVQALAGRADARTTAGDPDIAALEACLAEGERR-SLRDRISAH-FAL 398
QY 225 CMRPNOGYLIVQEMAQ 240
Db 399 -----GRAYLDLQDPAR 410

RESULT 12
Q82527 PRELIMINARY; PRT; 1093 AA.
AC Q82527;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Putative lysyl-tRNA synthetase.
GN Name=lys82; OrderedLocusNames=SAV7296;
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:526-531(2003).
CC -I- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
DR EMBL; AP005050; BAC75007.1; -.
DR HSSP; P14825; IE10.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0004815; F:aspartate-tRNA ligase activity; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004824; F:lysine-tRNA ligase activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0006422; F:aspartyl-tRNA aminoacylation; IEA.
DR GO; GO:0006430; F:lysyl-tRNA aminoacylation; IEA.
DR InterPro; IPR007424; DUF470.
DR InterPro; IPR007425; DUF471.
DR InterPro; IPR007426; DUF472.
DR InterPro; IPR008994; Nucleic_acid_OR.
DR InterPro; IPR004364; tRNA-synt_2.
DR InterPro; IPR002312; tRNA-synt_2.
DR InterPro; IPR002313; tRNA-synt_lys_2.
DR InterPro; IPR004365; tRNA anti.
DR InterPro; IPR006195; tRNA_ligase_II.
DR Pfam; PF04329; DUF470; 1.
DR Pfam; PF04330; DUF471; 1.
DR Pfam; PF04331; DUF472; 1.
DR Pfam; PF00152; tRNA-synt_2; 1.
DR Pfam; PF01336; tRNA anti; 1.
DR PRINTS; PR01042; TRNASYNTHASP.
```

```
DR PRINTS; PRO0982; TRNASYNTHLYS.
DR TIGRfams; TIGR00499; lys_bact; 1.
DR PROSITE; PS50862; AA_TRNA_LIGASE_II; 1.
KW ATP-binding; aminoacyl-tRNA synthetase; Complete proteome; Ligase;
KW Protein biosynthesis.
SQ SEQUENCE 1093 AA; 119034 MW; B916D8B13D6ABE53 CRC64;

Query Match 7.8%; Score 100; DB 2; Length 1093;
Best Local Similarity 26.5%; Pred. No. 26;
Matches 36; Conservative 15; Mismatches 47; Indels 38; Gaps 4;

QY 19 EALTENC-----REIFERRRHVLVGISPFNSRFSEDIYIRLIGWAKAQ 61
Db 939 EALHRRCDRAGVPYTTADGCGDWLEMYERLVEERTGLTFYKDFPTD----- 986
QY 62 FKSVSVLLAGHEA-ANLLE-----ALGTGPGKAERKVRKVEVSRNRRFAERLVAHGGDPK 115
Db 987 ----VSPLTRQHRADPRLAERWDLVAFGTGLTAYSELTDPMEQRRRLTAQSLLAAGDPE 1043
QY 116 AIHTSFDFIDNKAYQL 131
Db 1044 AMELDEDFDIALEYAM 1059

RESULT 13
Q980H7 PRELIMINARY; PRT; 392 AA.
AC Q980H7;
DT 01-OCT-2001 (TReMBLrel. 18, Created)
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN OrderedLocusNames=SSO0320;
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Awayez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Ebrau G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR EMBL; AS006666; NAK40655.1; -.
DR FIR; H90174; H90174.
DR InterPro; IPR008928; Glyco_trans_6bp.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 392 AA; 44603 MW; F7E9CBDDA820C556 CRC64;

Query Match 7.6%; Score 98.5; DB 2; Length 392;
Best Local Similarity 26.1%; Pred. No. 11;
Matches 65; Conservative 35; Mismatches 84; Indels 65; Gaps 14;

QY 8 VTERRSVHFIABALTENCRIE-ERRRHVLVGI-----SPFNSRFSEDIYIRLIGW 57
Db 83 VIESKDLKFIENKL-----RELLDEYKRIKVGKLGEGYTPDPIETPFSIIYETINKVIGG 138
QY 58 AKAQFKSVSVLL-----AGHEAANLLEALGTGPGKAERKVRKVEVSRN 99
Db 139 YEADTRMIEVLYTNTVYKEYLKAQKICSDDVARYLLE-----GKGKIKIDKSSAN 192
QY 100 RFAERLVAHGGDPKAIHTFSDPIDNKAYQLLRQVEHAPQPHRACLQMSREALI 159
Db 193 ----VSMVLNNG--VTSVKLLDFIDINSGVVRKKT-----KGILLD---EALA 236
QY 160 GRARGVSLMMEVSESDMLNAV---BYVIAELPFFIGAPDILEVEE-TLLAYHRPWKLG 215
```

```
Db 237 GNAL-VSEYRTINEEYLSAIVKADYIKNNLOHQRGRDKIENITKVPYLEP----- 290
QY 216 KISNHEPSI 224
Db 291 -ISNSEASI 298

RESULT 14
O75130
ID O75130 PRELIMINARY; PRT; 864 AA.
AC O75130
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DE K1AA0635 protein (Fragment).
GN Name=K1AA0635;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96112016; PubMed=8524414;
RA Wooster R., Bignell G., Lancaster J., Swift S., Seal S., Mangion J.,
RA Collins N., Gregory S., Gumbs C., Micklem G., Barfoot R., Hamoudi R.,
RA Patel S., Rice C., Biggs P., Hashim Y., Smith A., Connor F., Tonin P.,
RA Arason A., Gudmundsson J., Ficenec D., Kelsell D., Ford D., Tonin P.,
RA Bishop D.T., Spurr N.K., Ponder B.A.J., Seles R., Peto J., Devilee P.,
RA Corneliisse C., Lynch H., Narod S., Lenoir G., Egilsson V.,
RA Barkadottir R.B., Easton D.F., Bentley D.R., Futreal P.A.,
RA Ashworth A., Stratton M.R.;
RT "Identification of the breast cancer susceptibility gene BRCA2.";
RL Nature 378:789-792(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96172838; PubMed=8589730;
RA Tavtigian S.V., Simard J., Rommens J., Couch F., Shattuck-Eidens D.,
RA Neuhausen S., Merafiver S., Thorlacius S., Offit K., Stoppa-Lyonnet D.,
RA Belanger C., Bell R., Berry S., Bogden R., Chen Q., Davis T.,
RA Dumont M., Frye C., Hattier T., Jammulapati S., Janekci T., Jiang P.,
RA Kehr R., Leblanc J.-F., Mitchell J.T., McArthur-Morrison J.,
RA Nguyen K., Peng Y., Samson C., Schroeder M., Snyder S.C., Steele L.,
RA Stringfellow M., Stroup C., Swedlund B., Swensen J., Teng D.,
RA Thomas A., Tran T., Tran T., Tranchant M., Weaver-Feldhaus J.,
RA Wong A.K.C., Shizuya H., Eyfjord J.E., Cannon-Albright L., Labrie F.,
RA Skolnick M.H., Weber B., Kamb A., Goldar D.E.;
RT "The complete BRCA2 gene and mutations in chromosome 13q-linked
RT kindreds.";
RL Nat. Genet. 12:333-337(1996).
RN [3]
RP SEQUENCE FROM N.A., AND VARIANTS HIS-289; GLN-322; ASN-372; VAL-784;
RX SER-929; PHE-976; ILE-987; ASP-991; ASN-1561; LYS-1880; MET-1915;
RP PHE-2138; ARG-2462; ARG-2440; VAL-2466; THR-2490; PRO-2835; ALA-2856;
RP PHE-2944; THR-2951; ILE-3244 AND VAL-3412.
RA Rieder M.J., Livingston R.J., Daniels M.R., Chung M.-W.,
RA Miyamoto K.E., Nguyen C.P., Nguyen D.A., Poel C.L., Robertson P.D.,
RA Schackwitz W.S., Sherwood J.K., Witak L.A., Nickerson D.A.,
RT "NIHNS-SNPs, environmental genome project, NIHNS ES15478, Department
RT of Genome Sciences, Seattle, WA (URL: http://egp.gs.washington.edu.);
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RX PubMed=15057823; DOI=10.1038/nature02379;
RA Dunham A., Matthews L.H., Burton J., Ashurst J.L., Howe K.L.,
RA Ashcroft K.J., Beare D.M., Burford D.C., Hunt S.E.,
RA Griffiths-Jones S., Jones M.C., Keenan S.J., Oliver K., Scott C.E.,
RA Ainscough R., Almeida J.P., Ambrose K.D., Andrews D.T.,
RA Ashwell R.I.S., Babbage A.K., Baggeley C.L., Bailey J., Bannerjee R.,
RA Baylow K.F., Bates K., Beasley H., Bird C.P., Bray-Allen S.,
RA Brown A.J., Brown J.Y., Burrill W., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M.E., Clark S.Y., Clarke G., Clee C.M.,
RA Clegg S.C., Cobley V., Collins J.E., Corby N., Coville G.J.,
RA Deloukas P., Dhami P., Dunham I., Dunn M., Earthrowl M.E.,
RA Ellington A.G., Faulkner L., Frankish A.G., Frankland J., French L.,
RA Garner P., Garnett J., Gilbert J.G.N., Gilson C.J., Ghori J.,
RA Grafham D.V., Gribble S.M., Griffiths C., Hall R.E., Hammond S.,
RA Hart P.J., Hart E.A., Heath P.D., Howden P.J., Huckle E.J.,
RA Hunt P.J., Hunt A.R., Johnson C., Johnson D., Kay M., Kimberley A.M.,
RA King A., Laird G.K., Langford C.J., Lawlor S., Leongamornlert D.A.,
RA Lloyd D.M., Lloyd C., Loveland J.E., Lovell J., Martin S.,
RA Mashreghi-Mohammadi M., McLaren S.J., McMurray A., Milne S.,
RA Moore M.J.F., Nickerson T., Palmer S.A., Pearce A.V., Peck A.I.,
RA Pelan S., Phillimore B., Porter K.M., Rice C.M., Searle S.,
RA Sehra H.K., Shownkeen R., Skuce C.D., Smith M., Steward C.A.,
RA Sycamore N., Teeter J., Thomas D.W., Tracey A., Tromans A., Tubby B.,
RA Wall M., Wallis J.M., West A.P., Whitehead S.L., Willey D.L.,
RA Wilming L., Wray P.W., Wright M.W., Young L., Coulson A., Durbin R.,
RA Hubbard T., Sulston J.E., Beck S., Bentley D.R., Rogers J., Ross M.T.;
RT "The DNA sequence and analysis of human chromosome 13.";
RL Nature 428:522-528(2004).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 1519-1551 IN COMPLEX WITH
RP RAD51.
```

```
QY 2 NEWTGMVTERRSVHFIAEALTENCREFERRHVLVGISPFNS-----RFSYDIYRL- 54
Db 549 DDLATMARENQETISLELEAAVQEKEMKSRVHKVITEVSRESWESLMAAKEYENQDLDRFQ 608
QY 55 -----IGWAKAQKSVSVLLAGHEAANLEALGTPGKAERKVRKEVSNRRPFAE 104
Db 609 MLHNRADWEVKAHQAGSESSVRLE-----LLSIDTER----RHLRERVLELLEKEIQ 657
QY 105 RALVAHGDDPKAHTTFSDFDN--KAYQLLRQEVHEHAFPPQPHRHACLDMS--REALIG 160
Db 658 EHINAH-----HAYESQISSMAKAMSRLEELRHOEDEKATVFN---DLSSRLRELCK 707
QY 161 RARGVSLMVEESDMLNAVEYVIAELPFFIGAPDILEVE-----ETLLAYHR 209
Db 708 LDSGKDITMQLNSK--NLSEFVVVLELVNKSSEDLKQLNSRHTVTKNLSLLATNR 765
QY 210 PWKLGEKISNHFHSICMRPNQGYLIVGMAQLSEKRTITSE 250
Db 766 DKERFHSHTLSHE-----KDEIQLLKEKLT--LSSEKLTQS 799

RESULT 15
BRC2 HUMAN
ID BRC2 HUMAN STANDARD; PRT; 3418 AA.
AC P51587; O00183; O15008; Q13879;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Breast cancer type 2 susceptibility protein.
GN Name=BRC2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2004, 12:47:24 ; Search time 155.579 Seconds
(without alignments)
1497.801 Million cell updates/sec

Title: US-10-627-124-2
Perfect score: 2080
Sequence: 1 MSOSIKLPSVLSQFQNNPY.....AESGLYTRGPVSLVAFDGA 405

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt 02.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2080	100.0	405	1	CYPX_BACSU
2	958	46.1	407	2	Q7N9M6
3	825.5	39.7	408	2	Q83VE9
4	788.5	37.9	407	2	Q9RJQ7
5	497.5	23.9	404	2	Q73WL1
6	497.5	23.9	404	2	AAS04966
7	468	22.5	405	1	CPXK_SACER
8	455.5	21.9	412	2	Q70AS7
9	455.5	21.9	412	2	CAE53704
10	444	21.3	404	2	Q81Q07
11	444	21.3	404	2	AAT31748
12	442	21.2	404	2	Q6H86
13	438	21.1	404	2	Q73719
14	438	21.1	404	2	AAS41573
15	433.5	20.8	404	2	Q83X67
16	429.5	20.6	405	2	Q82E54
17	427	20.5	411	2	Q74437
18	427	20.5	411	2	AAS02661
19	424.5	20.4	400	1	C13C_XYLFA
20	424.5	20.4	424	2	Q93QD5
21	422.5	20.3	424	2	Q7BR29
22	422.5	20.3	424	2	AAR83738
23	418.5	20.1	396	1	YJIB_BACSU
24	417.5	20.1	410	2	Q737F3
25	417.5	20.1	410	2	AAS41609
26	417	20.0	410	2	Q9X5P9
27	415.5	20.0	397	2	Q59523
28	415.5	20.0	405	1	CPXM_BACSU
29	415	20.0	414	2	Q70AR9
30	415	20.0	414	2	CAE53712
31	414.5	19.9	399	1	C13C_XYLFT

32	414.5	19.9	433	2	Q89G11
33	414	19.9	411	2	Q735A2
34	414	19.9	411	2	AAS42160
35	413	19.9	398	2	Q70JZ2
36	413	19.9	398	2	CAE11256
37	412.5	19.8	408	2	Q7N397
38	409.5	19.7	410	1	CPXM_BACME
39	408.5	19.6	402	1	C123_MYCTU
40	407.5	19.6	410	2	Q73XE4
41	407.5	19.6	410	2	AAS04682
42	407	19.6	417	2	Q59910
43	404.5	19.4	402	1	C13B_XYLFT
44	400.5	19.3	410	1	CPXY_BACSU
45	400.5	19.3	433	2	Q9F2Q0

ALIGNMENTS

RESULT 1
CYPX_BACSU
ID_CYPX_BACSU STANDARD; PRT; 405 AA.
AC O34926;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Putative cytochrome P450 CYPX (EC 1.14.-.-)
GN Name=cypX; Synonyms=cypB, cyp134; OrderedLocusName=BSU35060;
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RA Lazarevic V., Soldo B., Rivolta C., Reynolds S., Maue C.,
RA Karamata D.;
RT "Nucleotide sequence of the 300-304 chromosomal segment of Bacillus
RT subtilis";
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377; DOI=10.1038/36786;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.-D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.-Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppe G., Guy B.J., Haga K., Haitech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M.,
RA Jones L.-M., Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M.,
RA Klein C., Kobayashi Y., Koetter P., Koningsstein G., Krogh S.,
RA Kumano M., Kurita K., Lapidus A., Lardinois S., Lauber J.,
RA Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Maue C.,
RA Medigue C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S.,
RA Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B.,
RA Park S.H., Parro V., Pohl T.M., Portetelle D., Porwollik S.,
RA Prescott A.M., Presecan E., Fujic P., Furnelle B., Rapoport G.,
RA Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B.,
RA Rose M., Sadaie Y., Sato T., Scanlan E., Schleich S., Schroeter R.,
RA Scoffone F., Sekiguchi J., Sekowska A., Seror S.J., Seror P.,
RA Shin B.S., Soldo B., Sorokin A., Tacconi E., Takagi T., Takahashi H.,
RA Takenaru K., Takeuchi M., Tamakoshi A., Tanaka T., Terpetra P.,
RA Tognoni A., Tosato V., Uchiyama S., Vandenbol M., Vannier F.,
RA Vassarotti A., Viari A., Wambutt R., Wambol M., Wedler H.,
RA Weitzenegger T., Winters P., Wipat A., Yamamoto H., Yamane K.,
RA Yasunoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumstein E.,
RA Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";

mutated
point

```

RL Nature 390:249-256(1997).
CC -I- SIMILARITY: Belongs to the cytochrome P450 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL; AF017113; AAC67280.1; -.
DR EMBL; Z99121; CAB15511.1; -.
DR PIR; F69611; F69611.
DR HSP; P33006; ICPT.
DR Subtilisin; BGI2580; cypX.
DR InterPro; IPR002397; BP450.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 2.
DR Pfam; PR00359; BP450.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Complete proteome; Heme; Hypothetical protein; Monooxygenase;
KW Oxidoreductase.
FT METAL 353 353 Iron (heme axial ligand) (By similarity).
SQ SEQUENCE 405 AA; 45473 MW; A943DDFEAB67BD01 CRC64;

Query Match 100.0%; Score 2080; DB 1; Length 405;
Best Local Similarity 100.0%; Pred. No. 1.1e-141;
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSQSIKLFVSLDQFNPNVAYFSQLEEDPVHYEESIDSYFISRYHDVRYILQHPDIFT 60
Db 1 MSQSIKLFVSLDQFNPNVAYFSQLEEDPVHYEESIDSYFISRYHDVRYILQHPDIFT 60
QY 61 TKSILVERAEVPMRGPVLAQMHGKHSKRIRVRSFTIGDLDHLSPLIKQNAENLLAPYL 120
Db 61 TKSILVERAEVPMRGPVLAQMHGKHSKRIRVRSFTIGDLDHLSPLIKQNAENLLAPYL 120
QY 121 ERGKSLVNDFGKTFACVCTMDMLGDKRDHEKISEHSGVADFTISISQSPPEARASLW 180
Db 121 ERGKSLVNDFGKTFACVCTMDMLGDKRDHEKISEHSGVADFTISISQSPPEARASLW 180
QY 181 CSQLSQYLMPIKERRVNPVSGDLISILCTSEYEGMALSKDILALINLVLLAATPADK 240
Db 181 CSQLSQYLMPIKERRVNPVSGDLISILCTSEYEGMALSKDILALINLVLLAATPADK 240
QY 241 TLALMYHLNNPEQNDVLAADSLVPRAIAETLRYPVQLIPROLSDQTVVGGMEIKK 300
Db 241 TLALMYHLNNPEQNDVLAADSLVPRAIAETLRYPVQLIPROLSDQTVVGGMEIKK 300
QY 301 DTIVFCMIGANRDPEAFQPDVFNHREDLGKISAFSGAARHLAFSGIHNCVGAFAK 360
Db 301 DTIVFCMIGANRDPEAFQPDVFNHREDLGKISAFSGAARHLAFSGIHNCVGAFAK 360
QY 361 NEIEIVANIVLDKRNIRLEEDFCYAESGLYTRGPVSLVAFDGA 405
Db 361 NEIEIVANIVLDKRNIRLEEDFCYAESGLYTRGPVSLVAFDGA 405

RESULT 2
Q7N9M6 PRELIMINARY; PRT; 407 AA.
ID Q7N9M6
AC Q7N9M6
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Similar to cytochrome P450.
GN OrderedLocNames=plu0296;
OS Photorhabdus luminescens (subsp. laumondii).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=141679;

```

```

RN SEQUENCE FROM N.A.
RP STRAIN=TT01;
RC MEDLINE=22957627; PubMed=14528314;
RA Duchaut E., Runiock C., Frangeul L., Buchrieser C., Givaudan A.,
RA Taouit S., Bocs S., Boursaux-Eude C., Chandelier M., Charles J.-F.,
RA Dassa E., Derose R., Derville S., Freyssinet G., Gaudriault S.,
RA Medigue C., Lanois A., Powell K., Sigquier P., Vincent R., Wingate V.,
RA Zouine M., Glaser P., Boemare N., Danchin A., Kunst F.,
RT "The genome sequence of the entomopathogenic bacterium Photorhabdus
RT luminescens."
RL Nat. Biotechnol. 21:1307-1313(2003).
CC -I- SIMILARITY: Belongs to the cytochrome P450 family.
DR EMBL; BX571859; CAE12591.1; -.
DR Photolista; plu0296; -.
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002397; BP450.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00359; BP450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN 1.
KW Complete proteome; Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 407 AA; 46297 MW; 7C3F545F86538D65 CRC64;

Query Match 46.1%; Score 958; DB 2; Length 407;
Best Local Similarity 46.6%; Pred. No. 9.2e-61;
Matches 185; Conservative 82; Mismatches 124; Indels 6; Gaps 3;

QY 8 FSVLSDFQFNPNVAYFSQLEEDPVHYEESIDSYFISRYHDVRYILQHPDIFTTKSILVER 67
Db 7 FNIHDFKFKNPDYDFDILHKQDLVYFEQSQNSYFYGKVEDVDAILK-SSIFNTKPLTAL 65
QY 68 AEPVWRGVPVLAQMHGKHSKRIRVRSFTIGDLDHLSPLIKQNAENLLAPYLERGKSD 126
Db 66 AEPVMDRVLVAQMEGSEHACKRKFIMQGLSRDVFNRYEPMIRKITEDLIQPYMEKGNID 125
QY 127 LVNDFGKTFAVCVCTMDMLGDKRDHEKISEHSGVADFTISISQSPPEARASLWSCFSOLS 186
Db 126 LVNDFGKTFAVCVCTMDMLGDKRDHEKISEHSGVADFTISISQSPPEARASLWSCFSOLS 185
QY 187 QYLMPIKERRVNPVSGDLISILCTSEYEGMALSKDILALINLVLLAATPADKTLALMI 246
Db 186 RLLKPIIDQRRNPVSKDIISIFC----QDTAMSMSEITACLNILAAATEPADKILAMML 241
QY 247 YHLLNPEQNDVLAADSLVPRAIAETLRYPVQLIPROLSDQTVVGGMEIKKDTIVFC 306
Db 242 NHLISNPMLDVVLKDRSLVRODAFEETRLTSPVQLIPREASEDVTISGIDIPKGAUVFC 301
QY 307 MIGAANRDPEAFQPDVFNHREDLGKISAFSGAARHLAFSGIHNCVGAAPAKNEIEIV 366
Db 302 MIGAANRDPEAFQPDVFNHREDLGKISAFSGAARHLAFSGIHNCVGAAPAKNEIEIV 361
QY 367 ANIVLDKRNIRLEEDFCYAESGLYTRGPVSLVAFD 403
Db 362 SNIIIDLLNLRPADHYHYQETGVYTRGPKSLLSFD 398

RESULT 3
Q83YE9 PRELIMINARY; PRT; 408 AA.
ID Q83YE9
AC Q83YE9
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Shy1.
GN Names=shy1;
OS Streptomyces hygroscopicus subsp. yingchengensis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=228732;
RN [1]
RP SEQUENCE FROM N.A.

```



```
RA Qin L., Wang H., Wu Z., Lu Y.T., Deng Z., Zhao G.;  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
CC -I- SIMILARITY: Belongs to the cytochrome P450 family.  
DR EMBL; AY260760; AAP21648.1; -.  
DR HSP; P33006; ICPT.  
DR GO; GO:0004497; F:monoxygenase activity; IEA.  
DR GO; GO:0006118; P:electron transport; IEA.  
DR InterPro; IPR002397; BP450.  
DR Pfam; PF00067; P450; 1.  
DR PRINTS; PR00359; BP450.  
DR PRINTS; PR00385; P450.  
KW Heme; Monoxygenase; Oxidoreductase.  
SQ SEQUENCE 408 AA; 44826 MW; B7137737217CDE05 CRC64;  
  
Query Match 39.7%; Score 825.5; DB 2; Length 408;  
Best Local Similarity 42.8%; Pred. No. 3.3e-51;  
Matches 173; Conservative 75; Mismatches 153; Indels 3; Gaps 3;  
  
QY 1 MSQSIKLFVLSQFQNNPYAFYSQLREEDPVHYEISDYFISRYHVDVRYILQHPDI-F 59  
DB 1 MSAAHLPDILSPFAANPYAVAVMREKEPLIWHTEATQSYIISRYEDVERVFKKAEF 60  
QY -60 TTSLVERAEPVNRGVPVLAQMKGHSKRIIVRSFIGDALDH-LSPLIKQNAENLLAP 118  
DB 61 TTDYNNWQLEPV-HGRTILQSGREHARRALVAPAFRGSDLSQKFLPVIERNRELIDA 119  
QY 119 YLBERGKSDLVNDGKTFVAVCTVMDMLGDKRDHEKISEWHSGVADFTITSQSPEARAHS 178  
DB 120 FRHTGSADIYNDVATRPVNVNVIADMLGLDKADHARFGHTVATVAFIGNLSGDEVAAG 179  
QY 179 LMCSEQLSQYLMPIKERRVNPVNSDLSILCTSEYEGMALSKDILALINVLAAATEPA 238  
DB 180 ERTVEFAEYMLPVIRERRANPGDILLSTLCAAEVDGVRMSDEDIKAFCSLLAAGGETT 239  
QY 239 DKTALMIYHLNPNQMDVLAADRSIVPRAIAETLYRKPVPOLIIPQLSQDTVVGMEI 298  
DB 240 DKAIGILANLLSHPDQAAVRAVDRSLIAPAAETLYRTPPVQIMRQSAFTDVEVTGGTI 299  
QY 299 KCDTIIVFCMIGANRDPEAFQPDVFNHREDIGKSAFSGAARHLAFSGIHCVCVGAAF 358  
DB 300 PAGATVTCLLIGANRDERRYDRDFDIPDDLTATTSFAAAGHLAFALGRHFCVGALL 359  
QY 359 AKNEIEIVANVLDKMNRIRLEEDFCVAESGLYTRGPVSLLVAF 402  
DB 360 AKAEVEGLNQLLDAMPDLALDGHDLVEQGVETRGKTLPVRF 403  
  
RESULT 4  
QYRJQ7 PRELIMINARY; PRT; 407 AA.  
AC QYRJQ7  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Putative cytochrome P450  
GN OrderedLocuNames=SCO0583; ORFNames=SCP55.07;  
OS Streptomyces coelicolor.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2) / M145;  
RX MEDLINE=21996410; PubMed=12000953; DOI=10.1038/417141a;  
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
RA Thomson N.R., James K.D., Harris D.B., Quail M.A., Kleser H.,  
RA Harper D., Batenan A., Brown S., Chandra G., Chen C.W., Collins M.,  
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
RA Huang C.-H., Kleser T., Lark L., Murphy L.D., Oliver K., O'Neil S.,  
RA Rabinowitsch E., Rajandream M.A., Rutherford K.M., Rutter S.,  
RA Seeger K., Saunders D., Sharp S., Squares R., Squares K., Taylor K.,  
RA Warren T., Wietzorrek A., Woodward J.R., Barrell B.G., Parkhill J.,
```

```
RA Hopwood D.A.;  
RT "Complete genome sequence of the model actinomycete Streptomyces  
RT coelicolor A3(2).";  
RL Nature 417:141-147(2002).  
CC -I- SIMILARITY: Belongs to the cytochrome P450 family.  
DR EMBL; AL939106; CAB61277.1; -.  
DR HSP; P33006; ICPT.  
DR GO; GO:0004497; F:monoxygenase activity; IEA.  
DR GO; GO:0006118; P:electron transport; IEA.  
DR InterPro; IPR002397; BP450.  
DR Pfam; PF00067; P450; 2.  
DR PRINTS; PR00359; BP450.  
DR PRINTS; PR00385; P450.  
KW Complete proteome; Heme; Monoxygenase; Oxidoreductase.  
SQ SEQUENCE 407 AA; 44916 MW; 86320B142522C788 CRC64;  
  
Query Match 37.9%; Score 788.5; DB 2; Length 407;  
Best Local Similarity 41.5%; Pred. No. 1.5e-48;  
Matches 169; Conservative 73; Mismatches 162; Indels 3; Gaps 3;  
  
QY 1 MSQSIKLFVLSQFQNNPYAFYSQLREEDPVHYEISDYFISRYHVDVRYILQHPDIF 59  
DB 1 MSTAQQVPDILSPFAANPYATVTRDSAPLIRHEATQSWIYSRYEDVERVFKDAGQF 60  
QY 60 TTSLVERAEPVNRGVPVLAQMKGHSKRIIVRSFIG-DALDHLSPLIKQNAENLLAP 118  
DB 61 TTDYNNWQLEPV-HGRTILQSGREHARRALVAPAFRGADLQERFLPVIERNRELIDA 119  
QY 119 YLBERGKSDLVNDGKTFVAVCTVMDMLGDKRDHEKISEWHSGVADFTITSQSPEARAHS 178  
DB 120 FRHTGRADLVADYATRPVNVNVIADMLGDKADHRRFGHTVATVAFIGNLSGDEVAAG 179  
QY 179 LMCSEQLSQYLMPIKERRVNPVNSDLSILCTSEYEGMALSKDILALINVLAAATEPA 238  
DB 180 ARTTFFAEYMLPVIRERRANPGDILLSTLCAAEVDGVRMSDEDIKAFCSLLAAGGETT 239  
QY 239 DKTALMIYHLNPNQMDVLAADRSIVPRAIAETLYRKPVPOLIIPQLSQDTVVGMEI 298  
DB 240 DKAIGIFANLLAHPQAAVRAVDRSLIAPAAETLYRTPPVQIMRQSAFTDVTLSGCTI 299  
QY 299 KCDTIIVFCMIGANRDPEAFQPDVFNHREDIGKSAFSGAARHLAFSGIHCVCVGAAF 358  
DB 300 PAGATVTCLLIGANRDERRYDRDFDIPDDLTATTSFAAAGHLAFALGRHFCVGALL 359  
QY 359 AKNEIEIVANVLDKMNRIRLEEDFCVAESGLYTRGPVSLLVAFDGA 405  
DB 360 AKAEVEIGVQLLDALPGLTDEGFEVVERGVETRGQSLPVRTTPA 406  
  
RESULT 5  
QY3WL1 PRELIMINARY; PRT; 404 AA.  
AC QY3WL1  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Hypothetical protein.  
GN OrderedLocuNames=MAP2649c;  
OS Mycobacterium paratuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;  
OC Mycobacterium avium complex (MAC).  
OX NCBI_TaxID=1770;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=k10;  
RA Li L., Bannantine J., Zhang Q., Amonsin A., Alt D., Kapur V.;  
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.  
CC -I- SIMILARITY: Belongs to the cytochrome P450 family.  
DR EMBL; AE017236; AAS04966.1; -.  
DR InterPro; IPR002397; BP450.  
DR InterPro; IPR001128; Cytochrome_P450.
```

DR	Pfam; PF00067; p450; 1.	
DR	PRINTS; PR00359; BP450.	
DR	PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN 1.	
KW	Complete proteome; Heme; Hypothetical protein; Monooxygenase; Oxidoreductase.	
SW	SEQUENCE 404 AA; 45054 MW; 3224F829F78A27FE CRC64;	
Query Match 23.9%; Score 497.5; DB 2; Length 404;		
Best Local Similarity 29.8%; Pred. No. 1.5e-27;		
Matches 119; Conservative 86; Mismatches 161; Indels 33; Gaps 7;		
QY	18 NPVAYFSQLEEDPVH-----YESIDSYFISRYHDVRYILOHPDIFTTKSLVERAE 69	
DB	24 DPTTELARLRREEPIQLRDTSGMPHEESKPVFIVYRHEDAQQMLRDNETSSAAVIAFG 83	
QY	70 PVMRGPVLAQMGKHEHSKRIVRVRSFIGDALDHL-SPLIKQNAENLLAPYLRGKSDLV 128	
DB	84 PVLGERVMLGIDEPVHGRLSLVSKAFSQKALARWEDELVGRVGNLSLIDRFAGNGKADLV 143	
QY	129 NDFGKTFACVVTMDMLGDKRDHEKISEHSGVADFTISQSPEARAHSLWCSEOLSOY 188	
DB	144 KEFTDPSRIIAGLGLPEQDPQFORWSISLLSWIL-----NPE---RGLAASAALCDY 196	
QY	189 LMPVIKERRVNPGSDLISILCTSEYEGMALSKDILALILNVLAAATEPADKTLALMIYH 248	
DB	197 FAPILAAARRAEPKDILLISGLAQAEIDGKLEDEEIIYFLRLLLPAGVETTYRALGSILLA 256	
QY	249 LLNPPQMDVLAQRSVLPRAETLRYKPPVQLIPRQLSQDVTVVGMEIKKDTIVFCMI 308	
DB	257 LLSDPQLDAIRGDSRLPQAIIEGVWRWEPPLLTITRVATRDTELGGVPIPGASTVMPML 316	
QY	309 GAANROPEAFEPQDVFNHREDLGKISAFSGAARHLAFSGIHCVCVGAFAAKNEIEIVAN 368	
DB	317 GAANROEDRYPPDRFDI-----FRAPKSHLGWGHVCLGMHLARLEMTAVN 366	
QY	369 IVLDMKRNIRLEEDFCYAES----GLYTRGPVSLLVAFD 403	
DB	367 LLLDRLPNLRLDPD---ADDPHIRGQVFRSPTSPVPLVD 402	
RESULT 7		
ID	CPXK_SACER STANDARD; PRT; 405 AA.	
AC	P33271;	
DT	01-FEB-1994 (Rel. 28, Created)	
DT	01-FEB-1994 (Rel. 28, Last sequence update)	
DT	05-JUL-2004 (Rel. 44, Last annotation update)	
DE	Cytochrome P450 107B1 (EC 1.14.-.-) (P450CYP107B1).	
GN	Names=CYP107B1;	
OS	Saccharopolyspora erythraea (Streptomyces erythraeus).	
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;	
OC	Pseudonocardineae; Pseudonocardaceae; Saccharopolyspora.	
OX	NCBI_TaxID=1836;	
RN	[1]	
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 172-203 AND 302-321.	
RC	STRAIN=NRRL 2338;	
RX	MEDLINE=92121109; PubMed=1732208;	
RA	Andersen J.F.; Hutchinson C.R.;	
RT	"Characterization of Saccharopolyspora erythraea cytochrome P-450	
RT	genes and enzymes, including 6-deoxyerythronolide B hydroxylase.;"	
RL	J. Bacteriol. 174:725-735(1992).	
CC	-!- FUNCTION: Not known, probably involved in the catabolism of octane	
CC	and guaiacol. It displays a weak activity in the O-dealkylation of	
CC	7-ethoxycoumarin.	
CC	-!- SIMILARITY: Belongs to the cytochrome P450 family.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL; M83110; AAA26483.1; -.	
DR	PIR; B42606; B42606.	
DR	HSSP; Q00441; 10XA.	
DR	InterPro; IPR002397; BP450.	
DR	InterPro; IPR001128; Cytochrome_P450.	
DR	Pfam; PF00067; p450; 2.	
DR	PRINTS; PR00359; BP450.	
DR	PRINTS; PR00385; P450.	
KW	Direct protein sequencing; Electron transport; Heme; Membrane;	
KW	Monooxygenase; Oxidoreductase.	
FT	METAL 352 352 Iron (heme axial ligand) (By similarity).	
SQ	SEQUENCE 405 AA; 45238 MW; 71C93CEC1FDC53FD CRC64;	
Query Match 22.5%; Score 468; DB 1; Length 405;		
Best Local Similarity 30.0%; Pred. No. 2e-25;		
Matches 115; Conservative 85; Mismatches 149; Indels 34; Gaps 8;		

```
QY 13 DOFQNNPYAFSOLREE--DPVHYBSIDSYFISRYHDVRYILOHP-----DIFTTKSLVE 66
Db 14 DAFQQRHNRKARMEEPQVIRTVNGLDQAWLITRYEDVKQALLDPRIAKDFGRTQOITE 73
QY 67 R--AEPVMR-----GPVLAQMHGKHSKRRIIVRSYFGDGLDHLSPILKQNAENLLA 117
Db 74 KRLADARRPGFSDIGPHMLNTDPDHTRLKLVVKAFARVEGLRPRIEQITDDLLD 133
QY 118 PYLERGKSLVNDFGKTFACVCTMDMLGDKRDHEKISEWHSGVADFPITISOSPEARAH 177
Db 134 RLAGESEVLDIDFAFPLPITVISLGMGVDSRRDDFRSNTWLVLD-----GSQPEAAQ 188
QY 178 SLWCSSQLSOYLMMPVIKERRVNPVPGSDLSILCTSEYEGMALSKOILALINVLAAATEP 237
Db 189 A---SVAMVEYLTELIAKRTKPEGDDLLTALLEAVEGDGRLSEGLIAMVFLLLVAGHET 245
QY 238 ADKTLALMIYHLNPNQMDVLAADRSVLRATAEILRYKPPV-QLIPQLSODTVVGGM 296
Db 246 TYNLIGNCVLSLGNPDQLAALRNDPSLLPGAIEETLRVESPVANGTFRHTAEAVRFGDV 305
QY 297 EIKKDTIVFCMIGANRDPFAFPQDVFNHREDLGIKSAFSGAARHLAFSGIHNCVGA 356
Db 306 VIEGELVWALGANRDERFEDPREDITRETTG-----HVAEGHGHFCVGA 355
QY 357 AFAKNEIEIVANVLDKRNIRL 379
Db 356 ALARLEAQIAGVRLRERFPDLRM 378
RESULT 8
Q70AS7
ID 070AS7 PRELIMINARY; PRT; 412 AA.
AC Q70AS7
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Putative cytochrome P450 reductase (EC 1.14.14.1).
GN Names=CYP0549;
OS Streptomyces peucetius.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1950;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 27952;
RA Parajuli N., Basnet D.B., Lee H.C., Sohng J.K., Liou K.;
RT "Genome analyses of Streptomyces peucetius ATCC 27952 for the
RT identification and comparison of cytochrome P450 complement with other
RT Streptomyces.";
RL Arch. Biochem. Biophys. 425:233-241(2004).
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
DR GO: GO:0004497; F:monooxygenase activity; IEA.
DR InterPro: IPR002397; BP450.
DR IntraPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450; 1.
DR PRINTS; PR00359; BP450.
DR PRINTS; PR00385; P450.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 412 AA; 45791 MW; 3F0D2529582C4F6A CRC64;
Query Match 21.9%; Score 455.5; DB 2; Length 412;
Best Local Similarity 29.3%; Pred. No. 1.7e-24;
Matches 125; Conservative 81; Mismatches 180; Indels 41; Gaps 12;
QY 1 MSQSIKFSVLSDFQFQNNPYAFSOLREE--DPVHYBSIDSYFISRYHDVRYILOHPDIFT 60
Db 1 MTQSL-LHQILDYANRADPPIVEELR-KTPVHHEED-GPYVVGTYEIRSLLDHPRISS 57
QY 61 -TKSLVERA-EPVWRG-----PVLQMHGKHSKRRIIVRSYFGDGLDHLSPILK 109
Db 58 DARNLASTAGDPLAESAQEEDSALPFSFLRLDPPHDLRMRNRSFGPPHSPHRVDGMR 117
QY 110 QNAENLLAPYL-----ERGKSLVNDFGKTFACVCTMDMLGDKRDHEKISEWHSGVADFI 165
Db 118 GEHLGIVSGLIDIGTGRIDLVEQSYFPPVTVICRLGVPREDEARFTWADTIA--- 174
QY 166 TSISQSPEA-----RAHSLWCSEOLSOYLMMPVIKERRVNPVPGSDLSILCTSEYEGMAL 218
Db 175 ASLDPNPDADPAERKASHD--ARTQLGMVLAGLIEERRKNPGDDILSELATAKGRDGTW 232
QY 219 SKDKILALINVLAAATEPADKTLALMIYHLNPNQMDVLAADRSVLRATAEILRYK 278
Db 58 DARNLASTAGDPLAESAQEEDSALPFSFLRLDPPHDLRMRNRSFGPPHSPHRVDGMR 117
```

```
QY 110 QNAENLLAPYL-----ERGKSLVNDFGKTFACVCTMDMLGDKRDHEKISEWHSGVADFI 165
Db 118 GEHLGIVSGLIDIGTGRIDLVEQSYFPPVTVICRLGVPREDEARFTWADTIA--- 174
QY 166 TSISQSPEA-----RAHSLWCSEOLSOYLMMPVIKERRVNPVPGSDLSILCTSEYEGMAL 218
Db 175 ASLDPNPDADPAERKASHD--ARTQLGMVLAGLIEERRKNPGDDILSELATAKGRDGTW 232
QY 219 SKDKILALINVLAAATEPADKTLALMIYHLNPNQMDVLAADRSVLRATAEILRYK 278
Db 233 TTMELLSTAALLIAGHETTIVNLVTNGMLTLRNPVQLRQRLADPRLAVPIVEELLRFEP 292
QY 279 PVOLIPRQLS-QDTVVGGMEIKDITVFCMIGANRDPFAFPQDVFNHREDLGIKSAF 337
Db 293 PVQLVPQRTTLTDIEVRGVTIPKASLWLVLASGNRDPQRFEDRFPDPRDI----- 346
QY 338 SGAARHLAFSGIHNCVGAFAKNEIEIVANVLDKRNIRLREDFCYAESSGLYTRGPVS 397
Db 347 ----QHLGSGSIHSCFGAPLARLEAHIALSELARLENPRLLDDPPPYRQNAVLRGPRH 402
QY 398 LIAVAFDG 404
Db 403 LPIACDG 409
RESULT 9
CAE53704
ID CAE53704 PRELIMINARY; PRT; 412 AA.
AC CAE53704;
DT 20-MAY-2004 (TrEMBLrel. 27, Created)
DT 20-MAY-2004 (TrEMBLrel. 27, Last sequence update)
DT 20-MAY-2004 (TrEMBLrel. 27, Last annotation update)
DE Putative cytochrome P450 reductase (EC 1.14.14.1).
GN CYP0549.
OS Streptomyces peucetius.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1950;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 27952;
RA Parajuli N., Basnet D.B., Lee H.C., Sohng J.K., Liou K.;
RT "Genome analyses of Streptomyces peucetius ATCC 27952 for the
RT identification and comparison of cytochrome P450 complement with other
RT Streptomyces.";
RL Arch. Biochem. Biophys. 425:233-241(2004).
DR ENBL; A605536; CAE53704.1; -.
KW Oxidoreductase.
SQ SEQUENCE 412 AA; 45791 MW; 3F0D2529582C4F6A CRC64;
Query Match 21.9%; Score 455.5; DB 2; Length 412;
Best Local Similarity 29.3%; Pred. No. 1.7e-24;
Matches 125; Conservative 81; Mismatches 180; Indels 41; Gaps 12;
QY 1 MSQSIKFSVLSDFQFQNNPYAFSOLREE--DPVHYBSIDSYFISRYHDVRYILOHPDIFT 60
Db 1 MTQSL-LHQILDYANRADPPIVEELR-KTPVHHEED-GPYVVGTYEIRSLLDHPRISS 57
QY 61 -TKSLVERA-EPVWRG-----PVLQMHGKHSKRRIIVRSYFGDGLDHLSPILK 109
Db 58 DARNLASTAGDPLAESAQEEDSALPFSFLRLDPPHDLRMRNRSFGPPHSPHRVDGMR 117
QY 110 QNAENLLAPYL-----ERGKSLVNDFGKTFACVCTMDMLGDKRDHEKISEWHSGVADFI 165
Db 118 GEHLGIVSGLIDIGTGRIDLVEQSYFPPVTVICRLGVPREDEARFTWADTIA--- 174
QY 166 TSISQSPEA-----RAHSLWCSEOLSOYLMMPVIKERRVNPVPGSDLSILCTSEYEGMAL 218
Db 175 ASLDPNPDADPAERKASHD--ARTQLGMVLAGLIEERRKNPGDDILSELATAKGRDGTW 232
QY 219 SKDKILALINVLAAATEPADKTLALMIYHLNPNQMDVLAADRSVLRATAEILRYK 278
Db 58 DARNLASTAGDPLAESAQEEDSALPFSFLRLDPPHDLRMRNRSFGPPHSPHRVDGMR 117
```

Db 233 TTWELLSTAALLIAGHETTWNLTNGMLTLLRNPDVLQRLRADPRLVPIVEBELLRFEP 292
QY 279 PVOLIPQLS-QTVVGMGMEIKDITVFCMIGANRDPEAFQPDVFNHIREDLGIKSAF 337
Db 293 PVQLVQRTTLLTIDIEVGRVTPKASLWULVLAGNRPQRFEDPDRPDORRDI----- 346
QY 338 SGAARHLAFSGHNCVGAFAKNEITEIVANIVLDKMRNIRLEEDFCYABSGLYTRGPVS 397
Db 347 ----QHLGLSGHSCFAGPLARLEAHIALSELARLENPRLLDPPYRQNAVLRGPRH 402
QY 398 LRVAFDG 404
Db 403 LPIACDG 409

RESULT 10
ID Q81Q07 PRELIMINARY; PRT; 404 AA.
AC Q81Q07; Q6HV76; Q6KS91;
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DT 01-OCT-2004 (T-EMBLrel. 28, Last annotation update)
DE Cytochrome P450 family protein.
GN OrderedLocusNames=BA2632, BAS2452; ORFNames=GBAA2632;
OS Bacillus anthracis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ames / isolate Porton;
RX MEDLINE=22608414; PubMed=12721629; DOI=10.1038/nature01586;
RA Read T.D., Peterson S.N., Tourasse N.J., Baillie L.W., Paulsen I.T.,
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA Holtzapple E.K., Ostad O.A., Helgason E., Rikstone J., Wu M.,
RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.L.,
RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA Benton J.L., Mahmoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA Hazen A., Cline R.T., Redmond C., Thwaitte J.E., White O.,
RA Salzberg S.L., Thomson B., Friedlander A.M., Kohler T.M.,
RA Hanna P.C., Kolstoe A.-B., Fraser C.M.;
RT "The genome sequence of Bacillus anthracis Ames and comparison to
RT closely related bacteria";
RL Nature 423:81-86(2003).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ames / isolate 0581;
RA Ravel J., Rasko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,
RA Wilson M., Stanley S., Decker S., Read T.D., Salzberg S.L.,
RA Fraser C.M.;
RT "Bacillus anthracis comparative genomics";
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Sterne;
RA Brettin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,
RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,
RA Richardson P., Rubin E., Tice H.;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the cytochrome P450 family.
DR EMBL; AE017032; AAP26480.1; -;
DR EMBL; AE017334; AAT31748.1; -;
DR EMBL; AE017225; AAT54763.1; -;
DR HSSP; Q55080; I107.
DR TIGR; BA2632; -;
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002397; BP450.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; p450; 2.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR00359; BP450.

DR PROSITE; PS00086; CYTOCHROME P450; UNKNOWN 1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 404 AA; 46728 MW; 960552E1D7792F55 CRC64;
Query Match 21.3%; Score 444; DB 2; Length 404;
Best Local Similarity 30.0%; Pred. No. 1.1e-23;
Matches 114; Conservative 68; Mismatches 176; Indels 22; Gaps 7;
QY 18 NPAYFSQREEDPVHYEESIDSYFISRYHDVRYILQHPDIFTTKSLIVERAE-----PVMR 73
Db 24 NPYEWYQFMRDNHPVHYDDEQDVWNVFLYDDVNRVLSYSLFSSRR--ERRQFAIPPLET 81
QY 74 GPVLAQMHGKHSKRIRIVRSFIGDALDHLSPKIKONAEHLAPYLERKCSLDVNDFGK 133
Db 82 RININSTDPPEHRNVRISVSKAFTPRSLQWKPRIQSIANELVKDIENCSEVDIVEQFAA 141
QY 134 TFAVCTVMDMLGLDKRDHEKISEWHSGVADFITSIQSPEARAHSLWCSEQLSQYLMFVI 193
Db 142 PLPVTVISDLGLGVPPTDRKKIKAW-SDILFMPYSEKFNLDLDAEKGIALNFKAYLLPIV 200
QY 194 KERRVNPNGSGLISILCTSEYEGMALSDKDLAILNLVLAATPADKTLALMIY-HLLNN 252
Db 201 QEKRYHLTDIIISDLIRAEYEGERLTDEEIVTFSGLLAAGNETTNNLIINSFYCFIVDS 260
QY 253 PEQMDVLADRSVAPRAIAETLRYKPPVOLIPQLSODTVVGGMEIKDITVFCMIGAN 312
Db 261 PATYKEVREKPKLISKAVEVLYRFPVTL-ARRITEDNIFGLPKKQDVWVWVSAAN 319
QY 313 RDEAFEPQDVFNHREDLGIKSAFGAARHLAFSGHNCVGAFAKNEIEIVANIVLD 372
Db 320 LDEKFSQASKFNHR-----IGNEKHLTFGKPHFCGLGAPLARLEAEIALTTFIN 370
QY 373 KRNIRILEDF-----CYAES 388
Db 371 AFEKIALSPFNIEQCILEN 390

RESULT 11
AAT31748 PRELIMINARY; PRT; 404 AA.
ID AAT31748 PRELIMINARY;
AC AAT31748;
DT 01-JUN-2004 (T-EMBLrel. 27, Created)
DT 01-JUN-2004 (T-EMBLrel. 27, Last sequence update)
DT 01-JUN-2004 (T-EMBLrel. 27, Last annotation update)
DE Cytochrome P450 family protein.
GN GBAA2632.
OS Bacillus anthracis str. Ames 0581.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;
OC Bacillus cereus Group; Bacillus anthracis.
OX NCBI_TaxID=261594;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ames 0581;
RA Ravel J., Rasko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,
RA Wilson M., Stanley S., Decker S., Read T.D., Salzberg S., Fraser C.M.;
RT "Bacillus anthracis comparative genomics";
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE017334; AAT31748.1; -;
SQ SEQUENCE 404 AA; 46728 MW; 960552E1D7792F55 CRC64;
Query Match 21.3%; Score 444; DB 2; Length 404;
Best Local Similarity 30.0%; Pred. No. 1.1e-23;
Matches 114; Conservative 68; Mismatches 176; Indels 22; Gaps 7;
QY 18 NPAYFSQREEDPVHYEESIDSYFISRYHDVRYILQHPDIFTTKSLIVERAE-----PVMR 73
Db 24 NPYEWYQFMRDNHPVHYDDEQDVWNVFLYDDVNRVLSYSLFSSRR--ERRQFAIPPLET 81
QY 74 GPVLAQMHGKHSKRIRIVRSFIGDALDHLSPKIKONAEHLAPYLERKCSLDVNDFGK 133
Db 82 RININSTDPPEHRNVRISVSKAFTPRSLQWKPRIQSIANELVKDIENCSEVDIVEQFAA 141
QY 134 TFAVCTVMDMLGLDKRDHEKISEWHSGVADFITSIQSPEARAHSLWCSEQLSQYLMFVI 193

```
Db 142 PLEVTWISDLGVPTTDRKKIKAW--SDILFMPYSKEKFNLDLDAEKGIALNEFKAYLLPIV 200
Qy 194 KERRVNPGLSILICTSEYEGMALSDKDLIALILNVLAAATPDKTTLALMIY-HLLNN 252
Db 201 QEKRYHLTDDIISDLIRAEYEGRLTDEEIVTSLGLAAGNETTNNLIINSFYCFLVDS 260
Qy 253 PEQNDVLADRSVLPRAIAETLRYPVQLIPROLSDQTVVGMGEIKKOTIVFCMIGAA 312
Db 261 PATYKEVREPKLISKAVEEVLRYRFPVTL-ARRITEDTNIIFGLMKDQMVAVWSAAN 319
Qy 313 RDEAFEPQDVFNHREDLGKSAFSGAARHLAFGSGIHNVCVGAFAKNEIEIVANIVLD 372
Db 320 LDEKKFSQASKFNHR-----IGNEKHLTFGKGFHCLGAPLARLEAEIALTTFIN 370
Qy 373 KMRNIRLEEDF-----CYAES 388
Db 371 AFEKIALSPSFNIEQCILEN 390
RESULT 12
Q6HI86 PRELIMINARY; PRT; 404 AA.
AC Q6HI86
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DE Cytochrome P450.
GN Name: cypA; ORFNames=BT9727.2414;
OS Bacillus thuringiensis serovar konkukian str. 97-27.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;
OC Bacillus thuringiensis serovar konkukian.
OX NCBI_TaxID=281309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=97-27;
RA Brettin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,
RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,
RA Richardson P., Rubin E., Tice H.;
RA Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the cytochrome P450 family.
DR EMBL; AE017355; AA160007.1; -.
DR InterPro; IPR002397; BP450.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00359; BP450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 404 AA; 46771 MW; C24B4C107972EFP4 CRC64;
Query Match 21.2%; Score 442; DB 2; Length 404;
Best Local Similarity 30.08; Pred. No. 1.5e-23;
Matches 114; Conservative 68; Mismatches 176; Indels 22; Gaps 7;
Qy 18 NPAYFSQLEEDPVHYEESIDSYFSRYHVDVRYILOHPDIFTTKSLVERAE-----PVMR 73
Db 24 NPWEYQFMRDNPVHYDDEQVNVFLYDQVNRVLSYRLFSRR--ERRQPAIPLET 81
Qy 74 GPVLAQMHGKHSANRIRVRSFPGDLDHLSPLIKQNAENLLAPYLERGKSLVNDFGK 133
Db 82 RININSTDPPEHNRVRSIVSKAFTPRSLQWKPRIOAIANELVQHIGKYSVNVESFAA 141
Qy 134 TFACVVTMDLGLDKRDHEKISEWHSGVADFTTSISQSPARAHSLWCSEQLSOYLMPIV 193
Db 142 PLPVTVISDLGLVPTTDRKKIKAW--SDILFMPYSKEKFNLDLDAEKGIALNEFKAYLLPIV 200
Qy 194 KERRVNPGLSILICTSEYEGMALSDKDLIALILNVLAAATPDKTTLALMIY-HLLNN 252
Db 201 QEKRYHLTDDIISDLIRAEYEGRLTDEEIVTSLGLAAGNETTNNLIINSFYCFLVDS 260
Qy 253 PEQNDVLADRSVLPRAIAETLRYPVQLIPROLSDQTVVGMGEIKKOTIVFCMIGAA 312
Db 261 PATYKEVREPKLISKAVEEVLRYRFPVTL-ARRITEDTNIIFGLMKDQMVAVWSAAN 319
Qy 313 RDEAFEPQDVFNHREDLGKSAFSGAARHLAFGSGIHNVCVGAFAKNEIEIVANIVLD 372
Db 320 LDEKKFSQASKFNHR-----IGNEKHLTFGKGFHCLGAPLARLEAEIALTTFIN 370
Qy 373 KMRNIRLEEDF-----CYAES 388
Db 371 AFEKIALSPSFNIEQCILEN 390
```

```
Qy 313 RDEAFEPQDVFNHREDLGKSAFSGAARHLAFGSGIHNVCVGAFAKNEIEIVANIVLD 372
Db 320 LDEKKFSQASKFNHR-----IGNEKHLTFGKGFHCLGAPLARLEAEIALTTFIN 370
Qy 373 KMRNIRLEEDF-----CYAES 388
Db 371 AFEKIALSPSFNIEQCILEN 390
RESULT 13
Q73719 PRELIMINARY; PRT; 404 AA.
AC Q73719
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DE Cytochrome P450 family protein.
GN OrderedLocusNames=BCE2659;
OS Bacillus cereus (Strain ATCC 10987).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=22523;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14960714;
RA Rasko D.A., Ravel J., Oekstad O.A., Helgason E., Cer R.Z., Jiang L.,
RA Shores K.A., Fouts D.E., Tourasse N.J., Angiuoli S.V., Kolonay J.F.,
RA Nelson W.C., Kolstoe A.-B., Fraser C.M., Read T.D.;
RT "The genome sequence of Bacillus cereus ATCC 10987 reveals metabolic
RT adaptations and a large plasmid related to Bacillus anthracis pXOI.";
CC -1- SIMILARITY: Belongs to the cytochrome P450 family.
DR EMBL; AE017272; AA541573.1; -.
DR TIGR; BCE2659; -.
DR InterPro; IPR002397; BP450.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00359; BP450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Complete proteome; Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 404 AA; 46830 MW; F0FD27FICD865640 CRC64;
Query Match 21.1%; Score 438; DB 2; Length 404;
Best Local Similarity 29.5%; Pred. No. 2.9e-23;
Matches 112; Conservative 73; Mismatches 173; Indels 22; Gaps 7;
Qy 18 NPAYFSQLEEDPVHYEESIDSYFSRYHVDVRYILOHPDIFTTKSLVERAE-----PVMR 73
Db 24 NPWEYQFMRDNPVHYDDEQVNVFLYDQVNRVLSYRLFSRR--ERRQFSIPLET 81
Qy 74 GPVLAQMHGKHSANRIRVRSFPGDLDHLSPLIKQNAENLLAPYLERGKSLVNDFGK 133
Db 82 RININSTDPPEHNRVRSIVSKAFTPRSLQWKPRIOAIANELVQHIGKYSVNVESFAA 141
Qy 134 TFACVVTMDLGLDKRDHEKISEWHSGVADFTTSISQSPARAHSLWCSEQLSOYLMPIV 193
Db 142 PLPVTVISDLGLVPTTDRKKIKAW--SDILFMPYSKEKFNLDLDAEKGIALNEFKAYLLPIV 200
Qy 194 KERRVNPGLSILICTSEYEGMALSDKDLIALILNVLAAATPDKTTLALMIY-HLLNN 252
Db 201 QEKRYHLTDDIISDLIRAEYEGRLTDEEIVTSLGLAAGNETTNNLIINSFYCFLVDS 260
Qy 253 PEQNDVLADRSVLPRAIAETLRYPVQLIPROLSDQTVVGMGEIKKOTIVFCMIGAA 312
Db 261 PGTYKEUREEPTLISKAIEEVLRYRFPITL-ARRITEDTNIIFGLMKDQMVAVWSAAN 319
Qy 313 RDEAFEPQDVFNHREDLGKSAFSGAARHLAFGSGIHNVCVGAFAKNEIEIVANIVLD 372
Db 320 LDEKKFSQASKFNHR-----IGNEKHLTFGKGFHCLGAPLARLEAEIALTTFIN 370
Qy 373 KMRNIRLEEDF-----CYAES 388
Db 371 AFEKIALSPSFNIEQCILEN 390
```

Db	371	AFBKIALSPSFNLEQCILEN	390	
RESULT 14				
AA541573		PRELIMINARY;	PRT;	404 AA.
AC	AA541573;			
DT	02-MAR-2004	(TrEMBLrel. 27, Created)		
DT	02-MAR-2004	(TrEMBLrel. 27, Last sequence update)		
DT	04-MAY-2004	(TrEMBLrel. 27, Last annotation update)		
DE	Cytochrome P450 family protein.			
GN	BCE2659.			
OS	Bacillus cereus (strain ATCC 10987).			
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
OX	NCBI_TaxID=222523;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	PubMed=14960714;			
RA	Rasko D.A., Ravel J., Oekstad O.A., Helgason E., Cer R.Z., Jiang L.,			
RA	Shores K.A., Fouts D.E., Tourasse N.J., Angiuoli S.V., Kolonay J.F.,			
RA	Nelson W.C., Kolstoe A.-B., Fraser C.M., Read T.D.,			
RT	"The genome sequence of Bacillus cereus ATCC 10987 reveals metabolic			
RT	adaptations and a large plasmid related to Bacillus anthracis pXOI."			
RL	Nucleic Acids Res. 32:977-988(2004).			
DR	ENBL; AE017272; AA541573.1; -.			
DR	TIGR; BCE2659; -.			
SQ	SEQUENCE 404 AA; 46830 MW; F0FD27F1CD865640 CRC64;			
Query Match	21.1%;	Score 438;	DB 2;	Length 404;
Best Local Similarity	29.5%;	Pred. No. 2.9e-23;		
Matches	112;	Conservative	73;	Mismatches 173; Indels 22; Gaps 7;
Qy	18	NPYAYSQLEEDPVHYEESIDSYFISRYHDVYILQHPDIFTKSLVERAE----	PVMR	73
Db	24	NPYEWYQFMRDNHPVHYDEBQDVNMVLYEDVARVLSYDLFSRR--ERRQFSIPLET	81	
Qy	74	GPVLAQMHGKHEGSAKRIRIVRSFTGDALDLSPLIKQNAENLLAPYLERGKSLDVLNDFGK	133	
Db	82	RIMNSTDPHEHNRVSIVSKATPFRSLEQWKPRIOAIANELVQHICKYSEVNIVEFAA	141	
Qy	134	TFAVCVTMGLDKRDHEKISEWHGSGVADPITSQSPEARAHSLWCSEQLSOYLMPIV	193	
Db	142	PLPVTVISDLGVPTTDRKKIKAW-SDILEWMPYSKEKFNLDVEKGIALNEFKAYLLPIV	200	
Qy	194	KERRVNPGLSILSTLCSTSEYEGWALSQDKLAILNLVLLAATPEADKTLALMIY-HLNN	252	
Db	201	QEKRYHLTDDIISLIRAEYEGKLTDEEIVTSLSGLAAGNETTNLIINSFYCFLVDS	260	
Qy	253	PEQMNVLADRSIVPRAIETLRYKPPVQLIPRQLSQDTVVGGMEIKKDTIVFCMIGAA	312	
Db	261	PGTYKELREPTLISKAIEVLRYRFPITL-ARRITEDTNI FGLMKKQDMVAVWSAAN	319	
Qy	313	RDEAFQPDVFNHREDLGKISAFSGAARHLAPGSGIHNCVGAAPAKNEIEIVANILVD	372	
Db	320	LDEKKSQASKFNHR-----IGNEKHLTFGKGFCLGAPLARLEAEIATLTFIN	370	
Qy	373	KMRNIRLEEDF----CYAES	388	
Db	371	AFBKIALSPSFNLEQCILEN	390	
RESULT 15				
Q83X67		PRELIMINARY;	PRT;	404 AA.
AC	Q83X67;			
DT	01-JUN-2003	(TrEMBLrel. 24, Created)		
DT	01-JUN-2003	(TrEMBLrel. 24, Last sequence update)		
DT	01-MAR-2004	(TrEMBLrel. 26, Last annotation update)		
DE	P450-like hydroxylase			
OS	Streptomyces rochei (Streptomyces parvullus).			
OG	Plasmid pSLA2-L.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Streptomycineae; Streptomycetaceae; Streptomyces.			
OX	NCBI_TaxID=1928;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=7434AN4;			
RX	MEDLINE=22676866; PubMed=12791134;			
RA	Mochizuki S., Hiratsu K., Suwa M., Ishii T., Sugino F., Yamada K.,			
RA	Kinaishi H.;			
RT	"The large linear plasmid pSLA2-L of Streptomyces rochei has an			
RT	unusually condensed gene organization for secondary metabolism."			
RL	Mol. Microbiol. 48:1501-1510(2003).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=7434AN4;			
RX	MEDLINE=20408175; PubMed=10954087;			
RA	Hiratsu K., Mochizuki S., Kinaishi H.;			
RT	"Cloning and analysis of the replication origin and the telomeres of			
RT	the large linear plasmid pSLA2-L in Streptomyces rochei."			
RL	Mol. Gen. Genet. 263:1015-1021(2000).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=7434AN4;			
RX	MEDLINE=20231737; PubMed=10767533;			
RA	Suwa M., Sugino H., Sasaoka A., Mori E., Fujii S., Shinkawa H.,			
RA	Nimi O., Kinaishi H.;			
RT	"Identification of two polyketide synthase gene clusters on the linear			
RT	plasmid pSLA2-L in Streptomyces rochei."			
RL	Gene 246:123-131(2000).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=7434AN4;			
RX	MEDLINE=99053144; PubMed=9836424;			
RA	Kinaishi H., Fujii S., Hatanai A., Kurokawa T., Shinkawa H.;			
RT	"Physical mapping of the linear plasmid pSLA2-L and localization of			
RT	the eryAI and actI homologs."			
RL	Biosci. Biotechnol. Biochem. 62:1892-1897(1998).			
CC	-1- SIMILARITY: Belongs to the cytochrome P450 family.			
DR	EMBL; AB088224; BAC76495.1; -.			
DR	HSSP; Q00441; 10XA.			
DR	GO; GO:0004497; F.monooxygenase activity; IEA.			
DR	GO; GO:0006118; P.electron transport; IEA.			
DR	InterPro; IPR002397; BP450.			
DR	InterPro; IPR001128; Cytochrome_P450.			
DR	Pfam; PF00067; P450; 1.			
DR	PRINTS; PR00359; BP450.			
DR	PRINTS; PR00385; P450.			
DR	PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN 1.			
KW	Heme; Monooxygenase; Oxidoreductase; Plasmid.			
SQ	SEQUENCE 404 AA; 45340 MW; IC9C00F7E7909891 CRC64;			
Query Match	20.8%;	Score 433.5;	DB 2;	Length 404;
Best Local Similarity	26.8%;	Pred. No. 6.2e-23;		
Matches	113;	Conservative	77;	Mismatches 169; Indels 63; Gaps 9;
Qy	12	SDQFQNNYAYSQLEEDPV-----HYESIDSYFISRYHDVYILQHP-----D	57	
Db	14	SELFHTDQATVATREILEQRFPVTRVRYDGLS-VMLVNRHEDVRAALTDPLSLNDPMKQSD	72	
Qy	58	I-----FTTKSLVERAEPVVRGPGVLAQMHGKHEGSAKRIRIVRSFTGDALDH	103	
Db	73	IDLASAATGIPADLIEYFQRMFRSDEP-----DHGRLKRLVTRFTVRRINA	119	
Qy	104	LSPLIKQNAENLLAPYLERGKSLDVLNDFGKTFACVVTMDMLGDKRDHEKISEWHSGVAD	163	
Db	120	LRPRIRQIADDDLEKFAATGGDLVEALARPLTWCELLGVPEEDRADFTQWSQHI--	177	
Qy	164	FITSISQSPEARAHSLWCSEQLSOYLMPIVKERRVNPGLSILSTLCSTSEYEGWALSQDI	223	
Db	178	-----VESSPEFAERNVAVSRSLFEVCVRSILRRRDEPGDLSALVDLRDVAORLSNEL	233	
Qy	224	LALILNVLAAATEPADKTLALMIYHLNNPEQMNVDLADRSIVPRAIETLRYKPPVQLI	283	
Db	234	ISTVFLVLAGIETTNNVLTGTGTFLLLTHPGELARLADGALLGPAVEEMLRVAPLEIT	293	

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2004, 13:25:54 ; Search time 30.2515 Seconds
(without alignments)
1288.126 Million cell updates/sec

Title: US-10-627-124-2
Perfect score: 2080
Sequence: 1 MSQSIXLFSVLSDQFQNNPY.....AESGLYTRGPVSLVAFDGA 405
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 segs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2080	100.0	405	1 F69611	cytochrome P450 cy
2	468	22.5	405	1 B42606	cytochrome P450 cy
3	424.5	20.4	400	2 E82817	cytochrome P-450 h
4	418.5	20.1	396	1 B69851	cytochrome P450 cy
5	415.5	20.0	397	1 S51594	cytochrome P450 m
6	409.5	19.7	410	2 S39924	cytochrome P450meg
7	408.5	19.6	402	1 A70707	cytochrome P450 rv
8	407	19.6	417	1 S49051	cytochrome P450 ty
9	400.5	19.3	410	1 B69611	cytochrome P450 cy
10	399.5	19.2	402	2 H82813	cytochrome P450-11
11	399.5	19.2	411	2 T36526	probable cytochrom
12	388.5	18.7	376	1 G69679	polyketide hydroxy
13	387	18.6	410	1 O4856M	cytochrome P450 bi
14	383.5	18.4	395	1 G69594	cytochrome P450 10
15	367.5	17.7	411	1 JC4003	cytochrome P450 -
16	359	17.3	412	1 B40634	erythromycin monoo
17	348.5	16.8	434	2 C87170	probable cytochrom
18	347.5	16.7	406	2 A35401	cytochrome P450 10
19	346.5	16.7	415	2 B97512	probable cytochrom
20	346.5	16.7	415	2 AC2731	cytochrome P450 by
21	345.5	16.6	408	1 S18531	cytochrome P450 er
22	343.5	16.5	428	1 F70729	cytochrome P450 rv
23	343	16.5	411	1 JC5859	polyketide synthas
24	338	16.2	398	1 F70807	cytochrome P450 rv
25	335	16.1	403	2 B35401	cytochrome P450 10
26	333.5	16.0	310	2 T44767	cytochrome P450 li
27	329	15.8	438	1 E70515	cytochrome P450 rv
28	328	15.8	407	2 B32306	cytochrome P450 cy
29	326.5	15.7	381	1 S15809	cytochrome P450 cy

make gene
day
teach
+
mily

RESULT 1
F69611
Cytochrome P450 CYPX - Bacillus subtilis
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Bacillus subtilis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
R:Accession: F69611
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero
C.; Bron, S.; Brouillette, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Evington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallerc
iedh, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.,
Kotter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Maeda, S.; Maueel,
Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle,
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, Y.; Sekowska, A.; Seror,
akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A. Bacillus subtilis.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: F69611
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-405 <KUN>
A:Cross-references: UNIPROT:O34926; GB:Z99121; GB:Z99122; GB:AL009126; NID:G2636029; PID:
A:Experimental source: strain 168
C:Genetics:
A:Gene: cypX
C:Superfamily: Bacillus cytochrome P450 CYP106 cytochrome P450 homology
C:Keywords: oxidoreductase
F:230-375/Domain: cytochrome P450 homology <CYP>

Query Match	100.0%	Score 2080;	DB 1;	Length 405;
Best Local Similarity	100.0%	Pred. No. 2.1e-149;		
Matches	405;	Conservative	0;	Mismatches
		Indels	0;	Gaps
				0;
QY	1	MSQSIXLFSVLSDQFQNNPYAFSQRLEDPVHYEESIDSYFISRVHYVYILQHPDIFT	60	
DB	1	MSQSIXLFSVLSDQFQNNPYAFSQRLEDPVHYEESIDSYFISRVHYVYILQHPDIFT	60	
QY	61	TKSLVERAEPVWRGVLQAQMGKHSKAKRIIVRSFIGDALDHLSPILKONENLAPYL	120	
DB	61	TKSLVERAEPVWRGVLQAQMGKHSKAKRIIVRSFIGDALDHLSPILKONENLAPYL	120	
QY	121	ERKGSDLVNDGKTFACVCTMDMLGDKRDHEKISEHWSGVADFTISISQSPARASLW	180	
DB	121	ERKGSDLVNDGKTFACVCTMDMLGDKRDHEKISEHWSGVADFTISISQSPARASLW	180	
QY	181	CSEQLSQYLMPIVKERRVNPVPGSDILISILCTSEYEGMALSDKOILAILNLVLAATPAK	240	

Db 181 CSEQLSQYLMPIKERRVNPGLDLSILCTSEYEGWALSDKDLILALINLVLAATEPADK 240

QY 241 TLALMIYHLNPNQNDVLADSLVPRAIAETRLKPPVQLIPQLSDQTVVGMWEIKK 300

Db 241 TLALMIYHLNPNQNDVLADSLVPRAIAETRLKPPVQLIPQLSDQTVVGMWEIKK 300

QY 301 DTVFCMIGAAANDPAPFQPDVFNHREDLIGKSAFSGAARHLAFSGGIHNCVGAAPAK 360

Db 301 DTVFCMIGAAANDPAPFQPDVFNHREDLIGKSAFSGAARHLAFSGGIHNCVGAAPAK 360

QY 361 NEIEIVANIVLDMKRNIRLEEDFCYAESGLYTRGPVSLVAFDGA 405

Db 361 NEIEIVANIVLDMKRNIRLEEDFCYAESGLYTRGPVSLVAFDGA 405

RESULT 2

B42606

Cytochrome P450 CYP1B1 - Saccharopolyspora erythraea

N:Contains: oxidoreductase (EC 1.-.-.-)

C:Species: Saccharopolyspora erythraea

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004

C:Accession: B42606

R:Andersen, J.F.; Hutchingson, C.R.

J. Bacteriol. 174, 725-735, 1992

A:Title: Characterization of Saccharopolyspora erythraea cytochrome P-450 genes and enzymology

A:Reference number: A42606; PMID:192121109; PMID:1732208

A:Accession: B42606

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA; protein

A:Residues: 1-405 <AND>

A:Cross-references: UNIPROT:P33271; GB:M83110; NID:gl52682; PIDN:AAA26483.1; PID:gl52684

A:Experimental source: NRRU2338

A:Note: sequence extracted from NCBI backbone (NCBIP:77484)

C:Genetics:

A:Gene: CYP107B1

C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology

C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase

F:238-374/Domain: cytochrome P450 homology <CYP>

F:352/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 22.5%; Score 468; DB 1; Length 405;

Best Local Similarity 30.0%; Pred. No. 1.1e-27;

Matches 115; Conservative 85; Mismatches 149; Indels 34; Gaps 8;

QY 13 DQFQNNPYAYFSQRLREE--DPVHYEESIDSYFISRVHVDVRYILQHP-----DIFTTKSLVE 66

Db 14 DAFAQDRHNRYARMREEPQVRIQTVNGLDAILTRYEDVKALLOPRIAKDFGRTOQIIE 73

QY 67 R--AEPVMR-----GPVLAQMHGKHSKRIVRSFIGDALDHLPLIKQNAENLLA 117

Db 74 KRLADARPGFSPDLGPHMLNTDPPDHLRLKLVVKAFTARRVEGLRPRISQITDLDLD 133

QY 118 PYLERKSLVNDVFGKTFACVCTMDMLGDKRDHEKISEWHSGVADFITSSQSPEARAH 177

Db 134 RLAGSEVDLIDEFAPPLPITWISLGMVEDSRDRDFTSWTNVLD-----GSQPEAQAO 188

QY 178 SLWCSSQLSQYLMPIKERRVNPGLDLSILCTSEYEGWALSDKDLILALINLVLAATEP 237

Db 189 A---SVAMVEYLTELLAKRTPEGDILLTALLEAVEDGDRLEGELIAMVFLLVAGHET 245

QY 238 ADKTLALMIYHLNPNQNDVLADSLVPRAIAETRLKPPV-QLIPQLSDQTVVGM 296

Db 246 TYNLIGNCVLSLGNPDQALARNDSLLPFGATEETLRYESPVANFTFRTAEAVRFGDV 305

QY 297 EIKKDTIVFCMIGAAANDPAPFQPDVFNHREDLIGKSAFSGAARHLAFSGGIHNCVGA 356

Db 306 VIPEGLVWVALGAANDGERFDDPRFDITRETTG-----HVAFGHGFVCGA 355

QY 357 AFANKEIEIVANIVLDMKRNIRL 379

Db 356 ALARLEAQIAVGRLLERFPDLRM 378

RESULT 3

E82817

Cytochrome P-450 hydroxylase XF0356 [imported] - Xylella fastidiosa (strain 9aSc)

C:Species: Xylella fastidiosa

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004

C:Accession: E82817

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequences

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; PMID:20365717; PMID:10910347

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: E82817

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-400 <SIM>

A:Cross-references: UNIPROT:Q9PGE6; GB:AE003887; GB:AE003849; NID:g9105175; PIDN:AAF8316

A:Experimental source: strain 9aSc

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Al-Brites, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carver, H. de-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohne J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.E.; Laigre chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.P.; Marino, C.L.; Marques, M.V.; Martins, E. A.; Authors: Martins, E.M.P.; Matukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Za

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF0356

C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology

F:235-370/Domain: cytochrome P450 homology <P45>

Query Match 20.4%; Score 424.5; DB 2; Length 400;

Best Local Similarity 28.0%; Pred. No. 2.1e-24;

Matches 111; Conservative 78; Mismatches 164; Indels 43; Gaps 6;

QY 5 IKLFSVLSDQFQNNPYAYFSQRLREEDPVHYEESIDSYFIS-----RYHDVRYILOH 55

Db 1 MKLADLSSPAFLENYPPLYETLRQ-----GSFVSIQGNALMTGYSIVDGLLHN 50

QY 56 PDIFTTKSLVERAE-----PVMRG--PVLQMGKHSKRIVRSFIGDALDH 103

Db 51 RNM--GKSYMESIRVRYGDDALDMLPFCQFNRMFLMLNPPVHTHLRGLVMQAFGRES 108

QY 104 LSPLIKQNAENLLAPYLERKSLVNDVFGKTFACVCTMDMLGDKRDHEKISEWHSGVAD 163

Db 109 MKPLATDTAHLRDDFEQKSSVDLTFEFPPLPFIICMMDDVDISDAISLVASVNIK 168

QY 164 FITTSQSPEARAHSLWCSSQLSQYLMPIKERRVNPGLDLSILCTSEYEGWALSDKDI 223

Db 169 VLDPAFMSFDELVHASAAVEELAHYFTRLIELRRAQPGTDLISMLLRAEEGQKLTHDEI 228

QY 224 LALIINLVLAATEPADKTLALMIYHLNPNQNDVLADSLVPRAIAETRLKPPVQLI 283

Db 229 VSNVTLLILGGVETTSNMIGNALIALHRHPKQARLKSLSLMPQAILCELRDYSVQFT 288

QY 284 PQLSDQTVVGMWEIKKTYIVFCMIGAAANDPAPFQPDVFNHREDLIGKSAFSGAARH 343

Db 289 MRAAMDDVSIQGVVFRGTIVFLMGAANRDPQFTDPPDHLDIRKQ-----GRL 338

QY 344 LAFSGGIHNCVGAAPAKNEIEIVANIVLDMKRNIRL 379

Db 339 QSGFAGVHHCLGYRLALVELECATVLLERLPHRL 374

RESULT 4

B69851

Cytochrome P450 yjib - Bacillus subtilis

N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Bacillus subtilis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: B69851
R:Kunze, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berteaux, C.; Bron, S.; Brouillet, S.; Bruch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choudhury, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabre, C.; Ferrarini, E.; Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallenbach, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Maeda, S.; Mausel, Y., M.; Osawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, A.; Schlegel, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron, A.; Tanaka, T.; Terpstra, P.; Toynoni, A.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Winters, P.; Wipat, A.; Yamamoto, H.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: B69851
A:Status: nucleic acid sequence not shown; translation not shown
A:Accession: B69851
A:Molecule type: DNA
A:Residues: 1-396 <KUN>
A:Cross-references: UNIPROT:O34374; GB:Z99110; GB:AL009126; NID:g2633472; PIDN:CAB13078.
A:Experimental source: strain 168
C:Genetics:
A:Gene: yliB
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C:Keywords: oxidoreductase
F:236-371/Domain: cytochrome P450 homology <CYP>
Query Match 20.1%; Score 418.5; DB 1; Length 396;
Best Local Similarity 26.2%; Pred. No. 5.8e-24;
Matches 95; Conservative 86; Mismatches 155; Indels 27; Gaps 6;
QY 18 NPAYFSQREEDPVYIESIDSYSFISRVHVDVYILQHPDFTTKSLVERAEPVWGPNVL 77
DB 25 HPPFWYEMRKAPVDFEENQVMSVFLYDDVKVGVGDKELFSS-CMPQOTSSI--GNSI 81
QY 78 AQMHGKHSKRIIVRSFTGDALHSLPKQNAENLLAPYLERGK-SDLVNDFGK 137
DB 82 INMDPPKTKIRSVNKAETPRVVKWQEPRIQBITDELQKQGRSEFDLVHDFSPVLPV 141
QY 138 CVTMDMLGLDKRDHEKISEHSGVADFTISQSPPEARAHSLW-----CSEQLSOYLM 191
DB 142 IVISELGVPSAHMEQFKW----SDLLVSTPKKSEAEKAFLEERDKCEELAAFFAG 197
QY 192 VIKERRVNPFGSDILISICTSEYEGMALSDKDLALILNVLAAATEPADKTLALMIYHLN 251
DB 198 IIEEKNKPEQDIISILVEAEETGKLSGBELIPFCTLLLVAGNETTTNLSNAMYSL 257
QY 252 NPEQMDVLADSLVPRATAETLYRPPVQLIPRQLSODTVVGGMEIKKDTIVFCMIGAA 311
DB 258 TPGVVEELSHPELMPQAVEEARFPAPVLRRIAKRDTTEIGHLIKGDMVLAFVAGA 317
QY 312 NRDPFAEPQDFVNIHREDLGKISAFSGAARHLAFSGHNCVGAFAFAKNEIHEIVANIVL 371
DB 318 NRDEAKFDRPHMFDIRH-----PNPHAFGHGHFCLGAPLARLE-----ANIAL 363
QY 372 DKM 374
DB 364 TSL 366
RESULT 5
S51594
N:Contains: Micromonospora griseorubida
C:Species: Micromonospora griseorubida
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: S51594
R:Inouye, M.; Takada, Y.; Muto, N.; Beppu, T.; Horinouchi, S.

Mol. Gen. Genet. 245, 456-464, 1994
A:Title: Characterization and expression of a P-450-like mycinamicin biosynthesis gene u
A:Reference number: S51593; MUID:95107242; PMID:7808395
A:Accession: S51594
A:Molecule type: DNA
A:Residues: 1-397 <INO>
A:Cross-references: UNIPROT:Q59523; EMBL:D16098; NID:g286050; PIDN:BAAO3672.1; PID:g3036
C:Genetics:
A:Gene: mycG
A:Start codon: GTG
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C:Keywords: heme; oxidoreductase
F:231-368/Domain: cytochrome P450 homology <CYP>
Query Match 20.0%; Score 415.5; DB 1; Length 397;
Best Local Similarity 28.9%; Pred. No. 9.8e-24;
Matches 114; Conservative 76; Mismatches 171; Indels 33; Gaps 11;
QY 23 FSQLEEDPVH-----YIESIDSYSFISRVHVDVYILQHPDFTTKSLVE---RAEPVMRG 74
DB 23 YGELQETEPVSRVPPYGE--EAWLVTRYEDVRAVLGDCGRFVRGPGSMTRDEPTRPEMVK 80
QY 75 PVLAQMHGKHSKRIIVRSFTGDALHSLPKQNAENLLAPYLERGK-SDLVNDFGK 133
DB 81 GGLSMDPPEHSLRLRLVVKATARRAESLRPRAREIAHELVDQMAATQPADLVAMFAR 140
QY 134 TFAVCVTMDMLGLDKRDHEKISEHSGVADFTISQSPPEARAHSLWSEQLSQYLMPIV 193
DB 141 QLPVRVICELGVPASADHDFRTRW-SGA--FLSTAETVAEMQEA--AEQAYAYMGDLI 194
QY 194 KERRVNPFGSDILISICTSEYEGMALSDKDLALILNVLAAATEPADKTLALMIYHLN 253
DB 195 DRRRKEPTDVLVSALVQARDQDLSLSEQLLGLVAGYESTTTQIADFFVYLLMTRP 254
QY 254 EQMDNDVLADSLVPRATAETLYRKP--PVQLIPRQLSODTVVGGMEIKKDTIVFCMIGAA 311
DB 255 EIRRQLDRPELIPSAVEELTRVPLGVTAPRVAVEDTVLRGVTIRAGEPVLASTGAA 314
QY 312 NRDPFAEPQDFVNIHREDLGKISAFSGAARHLAFSGHNCVGAFAFAKNEIHEIVANIVL 371
DB 315 NRDAQFPDADRIDVDRT-----PNQHLGFGHGHVHCLGAPLARVQLVALEVL 364
QY 372 DKMIRNL---BEDFCYAESGLYTRGPNVSLVAF 402
DB 365 QRLPGIRLIPETQLRWSE-GMLLRGPLEPVMV 397
RESULT 6
S39924
cytochrome P450meg - Bacillus megaterium (ATCC 13368)
C:Species: Bacillus megaterium
A:Variety: ATCC 13368
C:Date: 07-Oct-1994 #sequence_revision 23-Feb-1996 #text_change 09-Jul-2004
C:Accession: S39924; S32216
R:Rauchenbach, R.; Isernhagen, M.; Noeske-Jungblut, C.; Boidol, W.; Siewert, G.
Mol. Gen. Genet. 241, 170-176, 1993
A:Title: Cloning sequencing and expression of the gene for cytochrome P450meg, the stereo
A:Reference number: S39924; MUID:94049677; PMID:8232201
A:Accession: S39924
A:Molecule type: DNA
A:Residues: 1-410 <RAU>
A:Cross-references: UNIPROT:Q06069; EMBL:Z21972; NID:g288298; PIDN:CAA79985.1; PID:g28883
A:Experimental source: ATCC 13368
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C:Keywords: chromoprotein; heme; iron; metalloprotein
F:240-377/Domain: cytochrome P450 homology <P45>
F:355/Binding site: heme iron (Cys) (axial ligand) #status predicted
Query Match 19.7%; Score 409.5; DB 2; Length 410;
Best Local Similarity 26.8%; Pred. No. 2.9e-23;
Matches 109; Conservative 85; Mismatches 176; Indels 37; Gaps 9;
QY 3 QSIKFLSVLSQDFQNNPYAFYFSQLBEDPVHVEESIDSYFISRVHVDVYILQHPDIF--- 59

```

Db      8 KEITREKTRTEBF--SPYAWCKRMLENDPVSYHEGTDITWNVFYEDVKRVLSDYKHFSV 65
QY      60 ----TYSKSLVERAEPVNRGPV-----LAQMHGKEHSKRRIIVRSFIGDLDHLSPLIKON 111
Db      66 RKRTTISVGTDSB---EGSVPEKIQITSDPPDHRKRRSLLAFAFTFRSLQNWEPRIORI 122
QY      112 AENLLAPYLERGKSLVNDFGKTFACVVTMDMLGDKRDHEKISEWHSGVADFITSIQS 171
Db      123 ADELIGOMGGTEIDIVASLASPLPIIVWADLMGVPSKORLLFKW---VDTLFPFDFRE 179
QY      172 PEARAHSL--WCSEQSQYLMPVIKERRVNPVPGSDLSILCTSEYEGMALSKDKDILALILN 229
Db      180 KQEEVDKLAQVAKEYQYLYPIVVKRLNPADDIISDLKSEVDGEMFTDDVVRVTML 239
QY      230 VLLAATEPADKTLALMIYHLLNPNQMDVLADR-SLVPRATAETLYRYKPPVOLIPRQUS 288
Db      240 ILGAGVETTSHLLANSFYSLLYDKVEYQELHENLDLPQAVEEMLRFRFNLIKLDRTVK 299
QY      289 QTVVCGMEIKKDTIVFCMIGANRDPEAFQPDVFNHREDLGIKSAFSGAARHLAFGS 348
Db      300 EONDLLGVLEKGDVVVWMSAANDEEMFDPFTLNIHRPN-----NKKHLTFGN 350
QY      349 GHNCVGAFAFAKNEIIVANIVLDKRNIT-----RLEEDFCVAESG 389
Db      351 GPHFCLGAPLARLEAKIALTAFLKFKFHTAEVPSFQLEENLTDSATG 397

RESULT 7
A70707
Cytochrome P450 Rv0766c - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: A70707
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.T.; Davies, K.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
C:Accession: A70707
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-402 <COL>
A:Cross-references: UNIPROT:P77902; GB:Z80226; GB:AL123456; NID:g3261638; PIDN:CAB02396.
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: Rv0766c
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C:Keywords: oxidoreductase
F:236-372/Domain: cytochrome P450 homology <CYP>

Query Match          19.6%; Score 408.5; DB 1; Length 402;
Best Local Similarity 28.1%; Pred. No. 3.4e-23;
Matches 114; Conservative 73; Mismatches 172; Indels 47; Gaps 9;

QY      15 FQNNYAFPSQREEDPVHYEESIDSYFISRVHDVRYILQHPDIFTTKSLVERAEPVNRG 74
Db      19 FHEDPPYRRLRDEAPLYRNEERNFVAVSRHHDV---LQG---FRDSTALSNAVGVSLD 72
QY      75 P-----VLAQMHGKEHSKRRIIVRSFIGDLDHLSPLIKQNAENLLAPYLERG 123
Db      73 PSSRTSEAYRVMSMLAMDDPAHLRMTLVSKGTPRRIRELEPQVLELARIHLDLSALQTE 132
QY      124 KSLVNDFGKTFACVVTMDMLGDKRDHEKIS-----EWHSGVADFITSIQSPEARA 176
Db      133 SFDVFAEFAGKLPMDVISELIVGDTDRAKIRALADAVLHREDGVADV-----PPNAMA 186
QY      177 HSLWCSEQLSQYLMPVIKERRVNPVPGSDLSILCTSEYEGMALSKDKDILALILNVLLAATE 236
Db      187 ASI-----ELMRVYADLIAEFRPRPANNLTALLAAELDGDRLSDQISMAFLFLMWIAGNE 242
```

```

QY      237 PADKTLALMIYHLLNPNQMDVLADRSIVPRAIAETLYRYKPPVOLIPRQLSQDTPVVGW 296
Db      243 TTTKLLANAVWAAHHPGQLARVPADHSRIPMWVBETILRYDTSSQLARTVAHDLTYDT 302
QY      297 EIKKDTIVFCMIGANRDPEAFQPDVFNHREDLGIKSAFSGAARHLAFSGIHNVCVA 356
Db      303 TIPEGEVLLLLPGSANRRDRVDFDDPDDYRIGRE-IGCKL-----VSFGSAHFCLGA 353
QY      357 APAKNEIEIVANIVLDKRNIRLEEDFCVAESGLYTRG-----PVSL 398
Db      354 HILARMEARVALGALLRRIRNVEVDDNDNVVRVHSSNVVRGFAHLPISV 399

RESULT 8
S49051
Cytochrome P450 tyII (SC 1.1.-.-) [validated] - Streptomyces fradiae (strain T59235)
C:Species: Streptomyces fradiae
A:Variety: strain T59235
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: S49051
R;Merson-Davies, L.A.; Cundliffe, E. Mol. Microbiol. 13, 349-355, 1994
A:Title: Analysis of five tylosin biosynthetic genes from the tyIIA region of the Strept.
A:Reference number: S49051; MUID:95075319; PMID:7984112
A:Accession: S49051
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-417 <MER>
A:Cross-references: UNIPROT:Q59910; EMBL:U08223; NID:g6849140; PIDN:AAA21341.1; PID:g473;
A:Experimental source: strain T59235
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1994
C:Genetics:
A:Gene: tyII
C:Function:
A:Description: involved in C20ring oxidation of O-mycaminosyl tylactone [validated, MUID:
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C:Keywords: heme; oxidoreductase
F:253-388/Domain: cytochrome P450 homology <CYP>

Query Match          19.6%; Score 407; DB 1; Length 417;
Best Local Similarity 28.6%; Pred. No. 4.6e-23;
Matches 114; Conservative 78; Mismatches 163; Indels 44; Gaps 11;

QY      22 YFSQLREEDPVHYEESIDSYFISRVHDVRYILQHPDIFTTKSL-----VERAEPVNRGP-V 76
Db      38 WPARARAEAPVWDESROAQVFRYDDYLTVTSTNQLFSSDFSPVPVPEELAILMGPQT 97
QY      77 LAQMHGKEHSKRRIIVRSFIGDLDHLSPLIKQNAENLLAPYLERGKSLVDNDFGKTEA 136
Db      98 FGIDPPRHPGPKLVKSQAFTPRRTATLEPRIAETIRGLLDGRLKGQIDVVDLAYPLP 157
QY      137 VCVTMDMLGDKRDHEKISEWHSGVADFITSI--QSPPEARAHSLWCSEQLSQYLMPIVK 194
Db      158 VIVIAELLGIPAEDRDLPREW---VDVILNMEGEPN-----LPDDFSETMGPAIK 206
QY      195 E-----RRVNPVPGSDLSILCTSEYEGMALSKDKDILALILNVLLAATEPADKTLA 243
Db      207 EWGDYLYRRIALKRETPTDLDLSGLIEAVEGERRLTDEEIVNIVA-LLLTAGHISATLL 265
QY      244 LMIYHLLN-NPQMDVLADRSIVPRAIAETLYRYKPPVOLIPRQLSQDTPVVGMEIKKDT 302
Db      266 GMLFIVLDEHRAQAEELRADRLIFGAETELRYRSPFNIRFLLKEDTDLIGHPKAQG 325
QY      303 IYFCMIGANRDPEAFQPDVFNHREDLGIKSAFSGAARHLAFSGIHNVCVAFAKNE 362
Db      326 MVVAMIASANRDSAHFSDPDTEDVRRQ-----PNKMSFGHGHHCILGSLFLARLE 375
QY      363 IEIVANIVLDKRNIRLEEDFC--YAESGLYTRG-PVSL 398
Db      376 AKVFLLELFPDEFSDRYRVEHDEVEFEEDELTARRLPVTV 414
```

RESULT 9

E69611
Cytochrome P450 cypA - Bacillus subtilis
N;Contains: oxidoreductase (EC 1.-.-.-)
C;Species: Bacillus subtilis
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C;Accession: E69611; T44774
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berton
C.; Bron, S.; Brouillet, S.; Bruchsi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabbri, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallen
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
A.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9394377
A;Accession: E69611
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-410 <KUN>
A;Cross-references: UNIPROT:O08469; GB:Z99117; GB:AL009126; NID:g2634966; PIDN:CAB14615.
A;Experimental source: strain 168
R;Belitsky, B.R.; Gustafsson, M.C.U.; Sonenshein, A.L.; von Wachenfeldt, C.
J. Bacteriol. 179, 5448-5457, 1997
A;Title: An lrp-like gene of Bacillus subtilis involved in branched-chain amino acid tra
A;Reference number: Z22837; MUID:97431495; PMID:9287000
A;Accession: T44774
A;Status: translated from GB/ENBL/DBJ
A;Molecule type: DNA
A;Residues: 1-410 <BEL>
A;Cross-references: EMBL:Y11043; NID:g1926275; PIDN:CAA71937.1; PID:g1926278
A;Experimental source: strain 1A1
C;Genetics:
A;Gene: cypA
C;Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C;Keywords: oxidoreductase
F;245-381/Domain: cytochrome P450 homology <CYP>

Query Match 19.3%; Score 400.5; DB 1; Length 410;
Best Local Similarity 26.2%; Pred. No. 1.4e-22;
Matches 105; Conservative 95; Mismatches 162; Indels 39; Gaps 9;

QY 2 SQIKLFSVLSDQ-----FQNNPYAFSQLREEDPVHYBESIDS-----YFISRYHDVRY 51
DB 3 SKEKSVTILTESQLSSRAFKDEAYEFYKELKRSQAL-YPLSLGALGKGLISRYDDAIH 61
QY 52 ILQ-----HPDIFTTK-----SLVERAEPVMRGVLAQMHGKHSKRRIVRSFTGDA 100
DB 62 LLKNEKLKNVYENFTAKEKRPALLKNEETLTKH--MLNSDDPDHNRRLTLVOKAFTHRM 119
QY 101 LDHLSPLIKQNAENLAPYLERGKSLVNDFGKTFACVCTMDMLGLDKRHEKISEWHSG 160
DB 120 ILQLEDKIQIHADSLDLKQVKNPMNVLDVYAFPLPIIVISEMLGIPLEDQRQKFRVWSQA 179
QY 161 VADPITSISQSPARASHLWCSEQLSOYLMPIVKERRVNPVGSDLISILCTSEYEGMALSD 220
DB 180 IIDF-----SDAPERLOENDHLLGFEVYLSLVKRREREPAGDLISALIQASEGTQLST 235
QY 221 KQIALILNVLAAEPADKTLALMIYHLLNPNQMNVDLADSLVPRAIAETLRYKPPV 280
DB 236 EELYSMIMLLIVAGHETTNNLTNTMYALMCHHDQLEKLRQOQPDLMNSAIEEALRHPSPV 295
QY 281 QLP-RLQSDTVVGGMEIKKDTIVFCMIGANRDPPEAQQPDVFNTHREDLGKISAFSG 339
DB 296 ELTIRWTAEFFILHGOEIKRKDVIIITSLASANRDEKIFPNADIFDERKN----- 346

QY 340 AARHLAFSGSIHNCVGAFAAKNEIEIVANIVLDMNRNIRLE 380
DB -NRHIAFGHGNHFCGLGAQLARLEAKIAITLLRRCPCNIQLK 386

RESULT 10

H82813
Cytochrome P450-like enzyme XF0377 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C;Accession: H82813
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: H82813
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-402 <SIM>
A;Cross-references: UNIPROT:Q9PGCS; GB:AE003889; GB:AE003849; NID:g9105205; PIDN:AAF8318
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, M.C.; Franco, M.C.; Frohne
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krueger, J.E.; Kuramae, E.E.; Laigr
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Mence, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF0377
C;Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
F;235-370/Domain: cytochrome P450 homology <p45>

Query Match 19.2%; Score 399.5; DB 2; Length 402;
Best Local Similarity 27.7%; Pred. No. 1.6e-22;
Matches 110; Conservative 76; Mismatches 166; Indels 45; Gaps 6;

QY 5 IKLFSVLSDQFQNNPYAFSQLREEDPVHYEESIDSYFIS-----RYHDVRYILOH 55
DB 1 MKLTDLSNPAFLNPPLVETLRAQP-----FVSGINMALMGYSYLVDSLLHN 50
QY 56 PDFTTKSLVERAEPVMRGVLAQM-----HGKHSKRRIVRSFTGDA 102
DB 51 RNM--GKYMESMR-VRYGDSAADMLPFOAFSRMFTINPPAHTLRLGLVQAFTGRE 107
QY 103 HLSPLIKQNAENLAPYLERGKSLVNDFGKTFACVCTMDMLGLDKRHEKISEWHSGVA 162
DB 108 SMRPLAIDTAHQIIDLNFQKPSVLDVAEFAPFPFQMIQICOMDVIDGDAVTLGIAVAKIA 167
QY 163 DFITSISQSPARASHLWCSEQLSOYLMPIVKERRVNPVGSDLISILCTSEYEGMALSDK 222
DB 168 KVFDPSPMSADELVHASTAYEEELAQFTKUIERRTHPGTDLISLMFRAEDGEKLTDE 227
QY 223 ILALIINVLAAEPADKTLALMIYHLLNPNQMNVDLADSLVPRAIAETLRYKPPVQL 282
DB 228 IVSNVIMLLIAGYETTNNMIGNALIAHRRHQEQLALLKSLSLMPQAVSCLRYDGSVQF 287
QY 283 IPQLQSDTVVGGMEIKKDTIVFCMIGANRDPPEAQQPDVFNTHREDLGKISAFSGAAR 342
DB 288 TMAAMDIDIEVEGELVPRGTWVFLMGAANRDPQAQFTHPDQDITRKQ-----GR 337
QY 343 HLAFGSGSIHNCVGAFAAKNEIEIVANIVLDMNRNIRL 379
DB 338 LQSGFGAIHCHGLGYRLALIELECATTLTFLERLPHRL 374

RESULT 11

T34526

probable cytochrome P450 hydroxylase - Streptomyces coelicolor

C/Species: Streptomyces coelicolor

C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004

C/Accession: T36526

R/Murphy, L.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, May 1999

A/Reference number: Z21609

A/Accession: T36526

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-411 <MUR>

A/Cross-references: UNIPROT:Q9X8Q3; EMBL:AL049754; PIDN:CAB42023.1; GSPDB:GN00070; SCORE

A/Experimental source: strain A3(2)

C/Genetics:

A/Gene: SCODB:SCH10.14C

C/Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology

C/Keywords: heme; iron; metalloprotein

P;250-385/Domain: cytochrome P450 homology <P45>

F;363/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 19.2%; Score 399.5; DB 2; Length 411;

Best Local Similarity 27.7%; Pred. No. 1.7e-22;

Matches 106; Conservative 75; Mismatches 169; Indels 33; Gaps 7;

QY

15 FQNNPYAQSQRREDPVHVEISDSYFISYHVDVYILO-----HPDIFTT 61

Db

22 FVADPYPAFAELRARGRLYEPFDQWLVPFHADVSALLDRDLGRYQYRFTHEDFGRT 81

QY

62 KSLVERAEP--VMRGPVLAQMKGKHSKAKRIIVRSFIGDALHLSPLIKQNAENLLAPY 119

Db

82 PPPPEQ-EPHTTNDHGMLDLEPPDHTRIIRLKSFAFTPTVERLKPVYHGLADDLVARL 140

QY

120 LERKGSVLNDFGKTFAVCTMDMLGLDKRDHEKISEHWSGVADFITSISQSPEARAHL 179

Db

141 VAAGGGDLTDVAEPLPVAVIAEMLGIPESDRAPLPWSAICG-MYELNPSEETAAKAV 199

QY

180 WCSEQLSOYLMPIKERRVNPVSGDLISILCTSEYEGNALSQDKILA---LILNVLLAATE 236

Db

200 RASLDFSDYLRALIAARKEPFGDDLSGLIAAHDDDLRTEQEMISTCVLLLNAGHEATV 259

QY

237 PADKTLALMTYTHLLNPEQNDVLADSLVPRAETALRYKPPVQLIPROLSODTVVGGM 296

Db

260 NATNGWLAF--RHFDQALALRADHSLVPSAVEELMRDTPQLQFERWVLEIEIDGT 316

QY

297 EIKKDTIVFCMIGANRDPPEAFQPDVFNTHREDLGKISAFSGAARHAFSGSHNCVGA 356

Db

317 TLPRAEAVMLFGSANHPAVFTDPERLDLTRD-----NPHISFSAGIHYCIGA 366

RESULT 12

G69679

polyketide hydroxylase (EC 1.-.-.-) pksS - Bacillus subtilis

C/Species: Bacillus subtilis

C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004

C/Accession: G69679

R/Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berta

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch

A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle

iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.

Koester, P.; Koningsstein, G.; Kroch, S.; Kunano, M.; Kurita, K.; Lapidus, A.; Lardinois

A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauch

Y, M.; Ogawa, K.; Ogikawa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,

A/Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seroi

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yamamoto, K.; Yata, K.; Yoshida, K. A.; Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A/Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A/Reference number: A69580; MUID:98044033; PMID:9384377

A/Accession: G69679

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-376 <KUN>

A/Cross-references: UNIPROT:O31785; GB:Z99113; GB:AL009126; NID:G2634090; PIDN:CAB13607.1

A/Experimental source: strain 168

C/Genetics:

A/Gene: pksS

C/Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology

C/Keywords: oxidoreductase

F;240-376/Domain: cytochrome P450 homology <CYP>

Query Match 18.7%; Score 388.5; DB 1; Length 376;

Best Local Similarity 24.6%; Pred. No. 1e-21;

Matches 98; Conservative 80; Mismatches 161; Indels 59; Gaps 7;

QY 7 LFSVLSQFQNNPYAYFSQIRREDPVH-----FTTKSLVERAEPVMRGPVLAQMKGKHSKAKRIIVRSFI 97

Db 7 MFPHGKGFHHNPFVSLGRFREEPFHRFELKRFGATYPAWLITRYDDCMAFKDNRI 66

QY 48 DRYVILQHPDI-----FTTKSLVERAEPVMRGPVLAQMKGKHSKAKRIIVRSFI 97

Db 67 DVKNVNNQIKMLNVSEIDIFVSDHMLAKDTP-----DHTFLRSILVHOAFT 113

QY 98 GDALDHLSPILIKQNAENLLAPYLERGKSLVNDFGKTFVCTVMDMLGDKRDHEKISEW 157

Db 114 PRTIENLRGSIQIEAQLDEMKNKADIMKSFASPLPFIVISLWMLGPKEDRSQFOIW 173

QY 158 HSGVADPFTISQSPARASHLWCSEQLSOYLMPIKERRVNPVSGDLISILCTSEYEGMA 217

Db 174 TNAMVD-----TSEGNRELTNOALREFKDYIAKLHDDRRIKPKDDLISKLVAENGSK 227

QY 218 LSDKDLALILNLNLAATEPADKTLALMTYTHLLNPEQNDVLADSLVPRAETALRYK 277

Db 228 LSEKELYSLMLFLVAVAGLETTVNLGSGTLLALQHKKECKLQKQPEMIATAVEELLRYT 287

QY 278 PPVQLIP-RQLSODTVVGGMEIKKOTIVFCMIGANRDPPEAFQPDVFNHREDLGKISA 336

Db 288 SPVVMANRWAIEDFTYKSHIKRGDMIFIGISANRDPNFFENPEILINR----- 339

QY 337 FSGAARHLAFSGSHNCVGAAPAKNRIEIVANIVLDKM 374

Db 340 --SPNRHISFGFGIHFCLGAPLARLEGHIAFKAFAEEI 375

RESULT 13

O4BS6M

cytochrome P450 106 - Bacillus megaterium

N/Alternate names: cytochrome P450BM-1

N/Contains: oxidoreductase (EC 1.-.-.-)

C/Species: Bacillus megaterium

C/Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 09-Jul-2004

C/Accession: S07764; S17973

R/He, J.S.; Ruettinger, R.T.; Liu, H.M.; Fulco, A.J.

Biochim. Biophys. Acta 1009, 301-303, 1989

A/Title: Molecular cloning, coding nucleotides and the deduced amino acid sequence of P-

A/Reference number: S07764; MUID:90089408; PMID:2597681

A/Accession: S07764

A/Molecule type: DNA

A/Residues: 1-410 <HEJ1>

A/Cross-references: UNIPROT:P14762; EMBL:X16610; NID:G39626; PIDN:CAA34612.1; PID:G39627

A/Accession: S17973

A/Molecule type: protein

A/Residues: 1-25 <HEJ2>

C/Genetics:

A/Gene: CYP106

C/Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology

C/Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase;

F;1-410/Product: cytochrome P450 106 #status experimental <MAT>
F;241-378/Domain: cytochrome P450 homology <CYP>
F;356/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 18.68; Score 387; DB 1; Length 410;
Best Local Similarity 25.3%; Pred. No. 1.5e-21;
Matches 99; Conservative 81; Mismatches 173; Indels 38; Gaps 7;
QY 14 QFQNN-----PYAFYFQQLREEDPVHYEESIDSYFISRYHDVRYILOHPDIF-----TTK 62
DB 13 KFQSRAEFFPIQWYKEMLANSPVYFHEETNTWNVFOYEHVKQVLSNYDFSSDGQRTTI 72
QY 63 SLVERAEPVNRGPV--IAQMHGKEHSAKRIRVRSFPGDLDHLSPLIKQNAENLLAPYL 120
DB 73 FVGDSKKKSTSPITNLTPDHRKARSILLAAAFPRSLKNWEPRIKQIAADLVEAIQ 132
QY 121 ERKGSDLVNDFGKTFACVCTMDMLGLDKRDEKISEHSGVADFITISISP-----E 173
DB 133 KNSTINIVDLSGPFPSLVLIADLFGVPVKORYQFKW-----VDILFQVDQERLEE 184
QY 174 ARAHSLWCSPQLSOYLMPVIRKERRVNPGLSILICTSEYEGMALSDDILALILNLLA 233
DB 185 IEQEKQAGAEYQYLYPIVIEKESNLSDDIISDLQAEVDGETFTDEEIVHATMLLGA 244
QY 234 ATPADKTLALMYHLLNPEQM--NDVLRSLVPRAIETRLYKPPVQLIPQLSQDTV 292
DB 245 GVETTSIAINMFYSFLYDKSLYSELNRNRELAPKAVEMLRYRPHISRDRTVKQDNE 304
QY 293 VGMGEIKKDTIVFCMIGAAANDPEAFQPDVFNHREDLGKSAFSAARHLAFSGSIHN 352
DB 305 LLGVKLLKGGDVIAWMSACNMDETFFENPFPSVDLHRET-----NKKHLTFGNGPHF 355
QY 353 CVGAFAAKNEIEIVANIVLDKMRNIRLEEDF 383
DB 356 CLGAPLARLEMKLLEAFLEAFSHIEPFEDF 386

RESULT 14

G69594
cytochrome P450 bioI - Bacillus subtilis
N/Contains: oxidoreductase (EC 1.-.-.-)
C/Species: Bacillus subtilis
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C/Accession: G69594
R;Kunst, F.; Ogawara, N.; Moberg, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
Y., M.; Ogawa, K.; Ogawara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A/Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A/Authors: Yoshikawa, H.P.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A/Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A/Reference number: A69580; MUID:98044033; PMID:9384377
A/Accession: G69594
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-395 <KUN>
A/Cross-references: UNIPROT:P53554; GB:Z99119; GB:AL009126; NID:92635411; PIDN:CAB14997.
A/Experimental source: strain 168
C/Genetics:
A/Gene: bioI
C/Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C/Keywords: oxidoreductase
F;232-367/Domain: cytochrome P450 homology <CYP>

Query Match 18.4%; Score 383.5; DB 1; Length 395;

Best Local Similarity 25.7%; Pred. No. 2.5e-21;
Matches 98; Conservative 87; Mismatches 175; Indels 21; Gaps 6;
QY 9 SVLSDFQNNPYAFYFQQLREEDPVHYEESIDSY--FISRYHDVRYILOHPDIFTTKSLV 65
DB 6 STASSFELKMPYSFYDTLRAVFI--YKSPFLKYPGWYVYGYEETAAALKDARFKVPTPLP 64
QY 66 ERAEPV-----MRGPVLAQMHGKEHSAKRIRVRSFPGDLDHLSPLIKQNAENLLAPYL 120
DB 65 ESSTKYQDLSHVQNMWLFQNPQDHRRLRLTLAGSAGTPRTTESYQPYIIEIVHLLDQVQ 124
QY 121 ERKGSDLVNDFGKTFACVCTMDMLGLDKRDEKISEHSGVADFITISISPSPARASLW 180
DB 125 GKXMEVISDFAPFLASFVIANIIGVPEDEQLEKWAASLIQTI--DFTSRKALTGNI 183
QY 181 CSQLSQYLMPVIRKERRVNPGLSILICTSEYEGMALSDDILALILNLLAATEPADK 240
DB 184 MAVQMAVYFELQKRRHPQODMISMLLGR--EKDKLTEEBAASTCILLAIAGHETVN 242
QY 241 TLALMYHLLNPEQMNDVLRSLVPRAIETRLYKPPVQLIPQLSQDTVVGMGIKK 300
DB 243 LISNVLCLLQHPQLKRENPDILGTAVEELRYESPQTOMTARVASEDIDICGVITRQ 302
QY 301 DTIVFCMIGAAANDPEAFQPDVFNHREDLGKSAFSAARHLAFSGHNCVGAAPAF 360
DB 303 GEQVYLLGAAANDPSIFTNPVDVDIR-----SPNHLSPFGHHVCLGSLAR 352
QY 361 NEIEIVANIVLDKMRNIRLEE 381
DB 353 LEAQIAINTLLQRMPSNLAD 373

RESULT 15

JC4003
cytochrome P450 - Streptomyces sp.
N/Contains: oxidoreductase (EC 1.-.-.-)
C/Species: Streptomyces sp.
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C/Accession: JC4003
R;Arisawa, A.; Tsunekawa, H.; Okamura, K.; Okamoto, R.
Biosci. Biotechnol. Biochem. 59, 582-588, 1995
A/Title: Nucleotide sequence analysis of the carbamycin biosynthetic genes including the
A/Reference number: JC4001; MUID:95290751; PMID:7772821
A/Accession: JC4003
A/Molecule type: DNA
A/Residues: 1-411 <ARI>
A/Cross-references: DDBJ:D30759; NID:9551628; PIDN:BA06420.1; PID:9551630
C/Note: the source was designated as Streptomyces thermotolerans
C/Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C/Keywords: oxidoreductase
F;242-382/Domain: cytochrome P450 homology <CYP>

Query Match 17.7%; Score 367.5; DB 1; Length 411;
Best Local Similarity 27.0%; Pred. No. 4.3e-20;
Matches 108; Conservative 84; Mismatches 161; Indels 47; Gaps 12;
QY 9 SVLSDFQNNPYAFYFQQLREEDPV---HYEESIDSYFISRYHDVRYILOHPDIFTTKSLV 65
DB 5 NLMDEPVRDPFGAYAEIRQAQELVPRASYPWGAVQWLATRYSDKAVLTPRLVNNPANV 64
QY 66 -----ERA-----EPV--MRGPVLAQMHGKEHSAKRIRVRSFPGDLDHLS 106
DB 65 PEMHLPHPYEQALGDGGIPDEYVRYLAGSILSQ--DGPGLRLRLRLSSRAFTVRRVNALRP 123
QY 107 LIQKNAENLLAPYLER---GKSLVNDVFGKTFACVCTMDMLGLDKRDEKISEHSGVAD 163
DB 124 RVTELARHLASLPDRAQDGVLDVLEDFSYPLSIDVICEIVGPEEARE---QWHTWGA 180
QY 164 FITSISQ--SPEARAHSLWCSEOLSOYLMPVIRKERRVNPGLSILICTSEYE--GMALSD 220
DB 181 FYTMDPAVITGPAVRG-----MADHLHLIIQRRATPTGDLTLGLVQAEDEGEPLTD 232
QY 221 KDILALILNLLAATEPADKTLALMIYHLLNPEQMNDVLRSLVPRAIETRLYKPPV 280

Db 233 EEIATLVITFTAGNETTAHLIGNGVAALITHSDQLALLRSDRRLLSQAVDELMRWCTPV 292
Qy 281 QLI-PROLSODTVVGGMEIKKDTIVFCMIGAAARDPEAFQDPDVFNHREDLGKSAFSG 339
Db 293 QVTQPRYATEDLDVGGVTVRKGEQVAVI GAAGHD PDRFPDPERFDITRNHRAPHEA--- 349
Qy 340 AARHLAFSGIHNVCVGAFAKNEIEIVANIVLDKMRNRL 379
Db 350 ---HVGFGFGPHYCLGAALAHQETALDITLDRFPFSLAL 386

Search completed: December 10, 2004, 13:45:49
Job time : 31.2515 secs